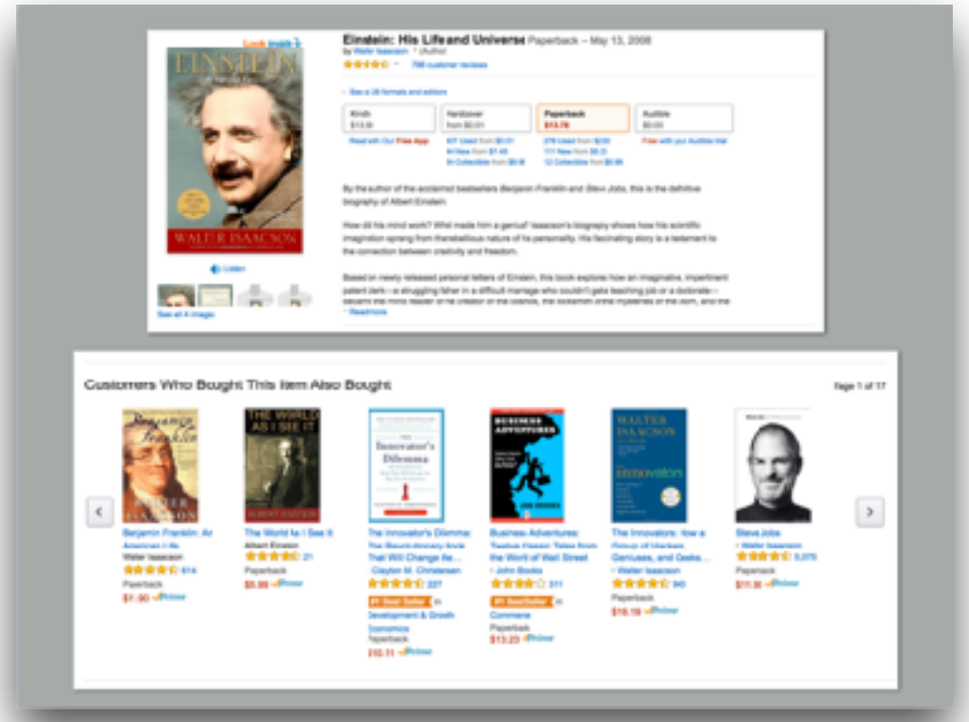
Representação multimodal em aprendizado de máquina: interpretabilidade e *few-shot learning*

Anisio Mendes Lacerda

Apresentado como requisito parcial para o concurso
de professor Adjunto Nível 1 - DCC/UFMG

Aprendizado de máquina

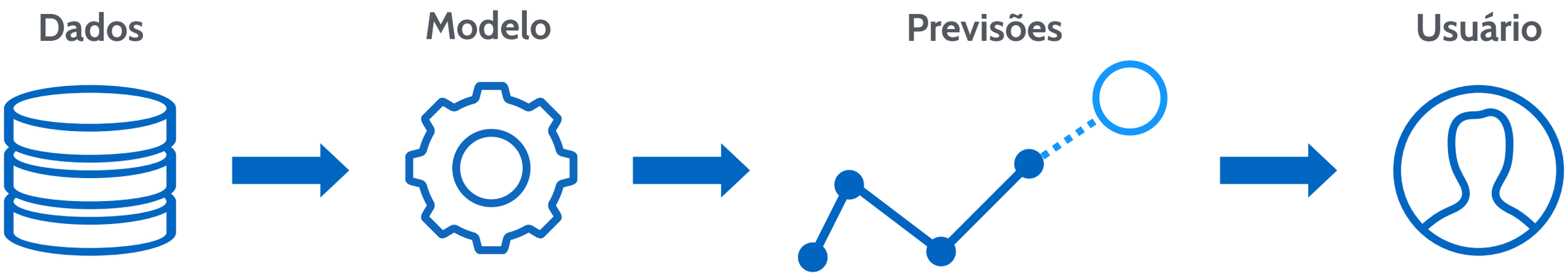
Sistemas de recomendação



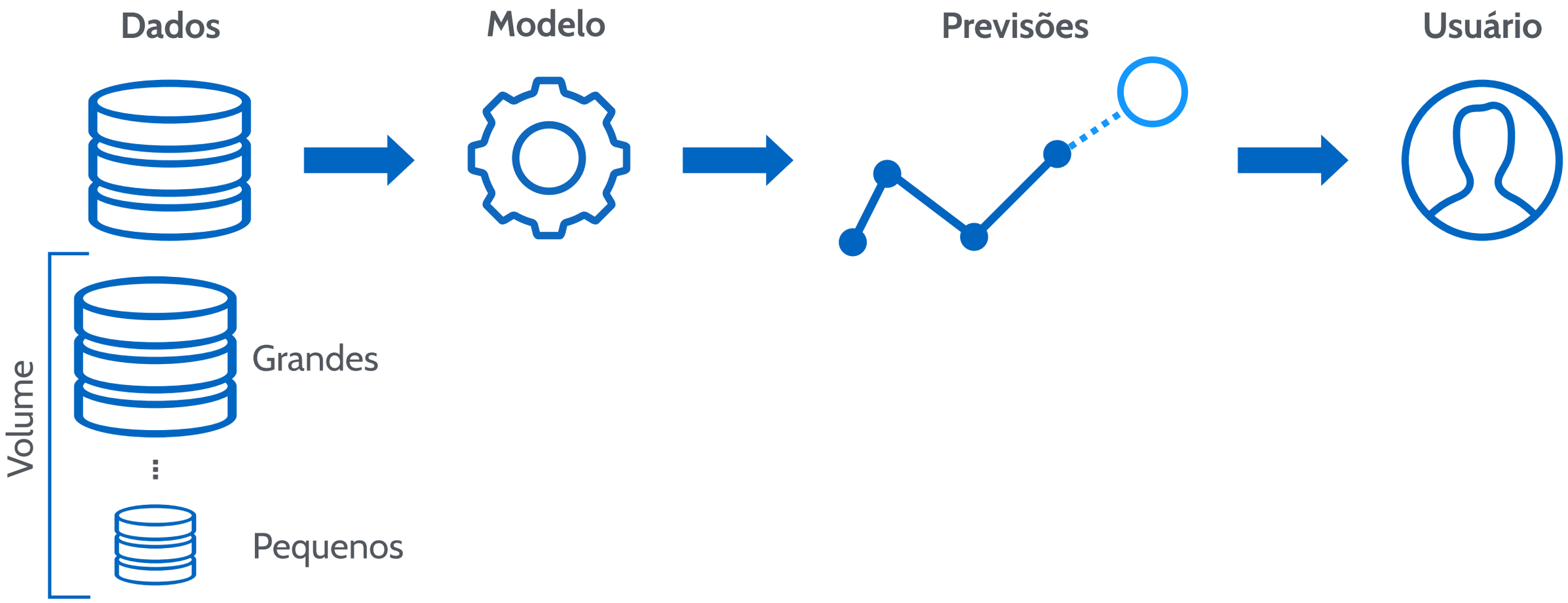
Filtros

Aplicações médicas

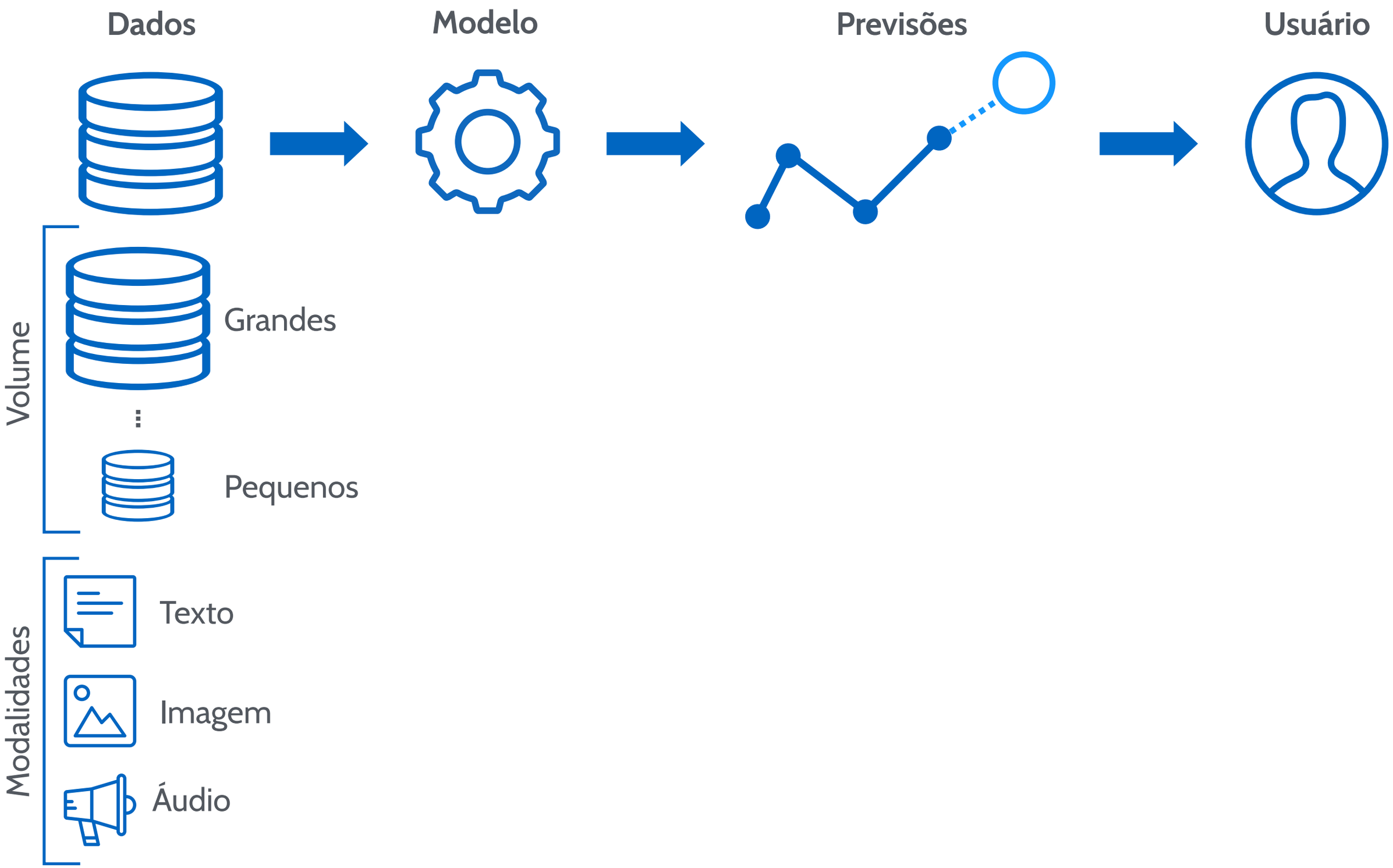
Contextualização



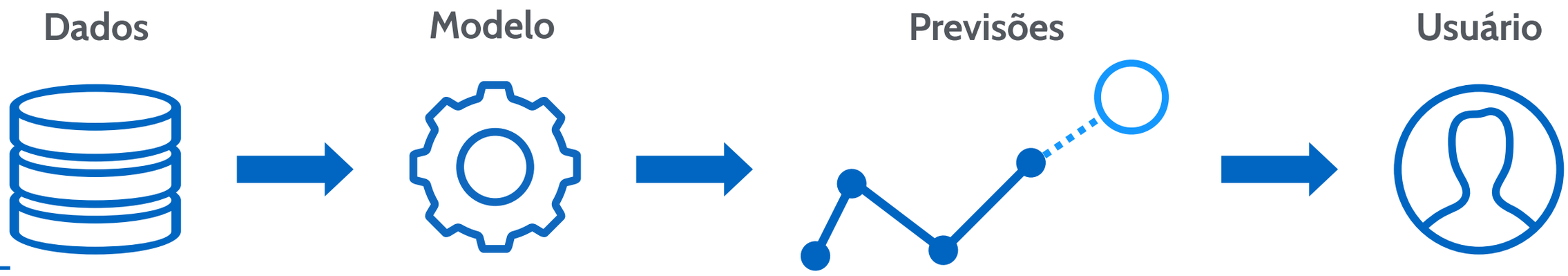
Contextualização



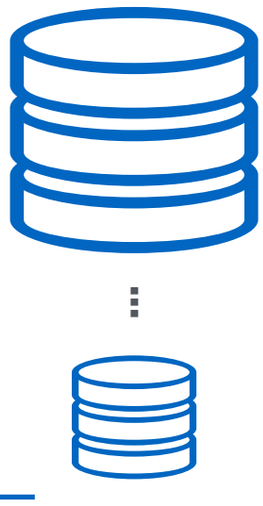
Contextualização



Contextualização



Volume



Grandes

Pequenos

Modalidades

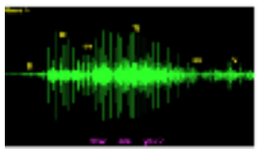
Texto

Recuperação de Informação

Imagem

Visão computacional

Áudio



Análise do discurso

Contextualização

Dados

Modelo

Previsões

Usuário

Volume

Grandes

Pequenos

Modalidades

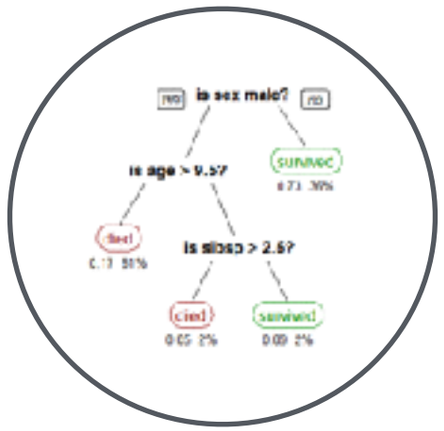
Texto

Imagem

Áudio

Tipos

Interpretáveis



Não-interpretáveis

Contextualização

Dados

Modelo

Previsões

Usuário

Volume

Grandes

Pequenos

Modalidades

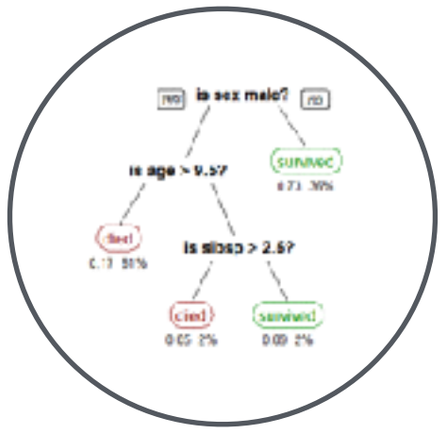
Texto

Imagem

Áudio

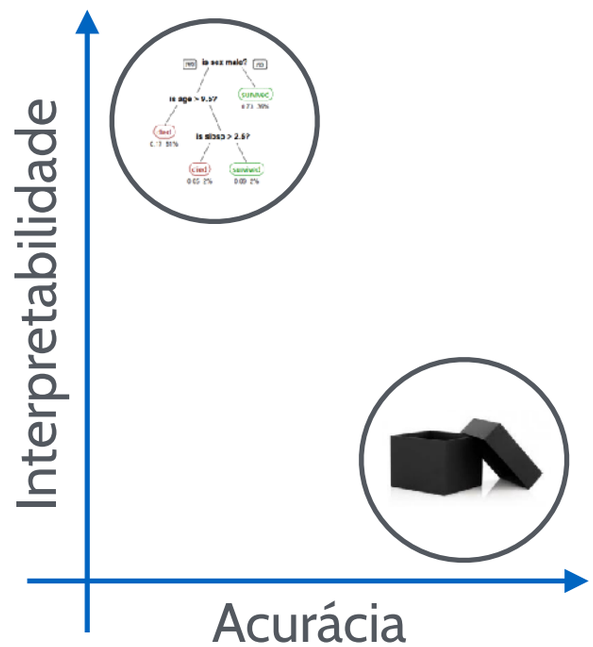
Tipos

Interpretáveis

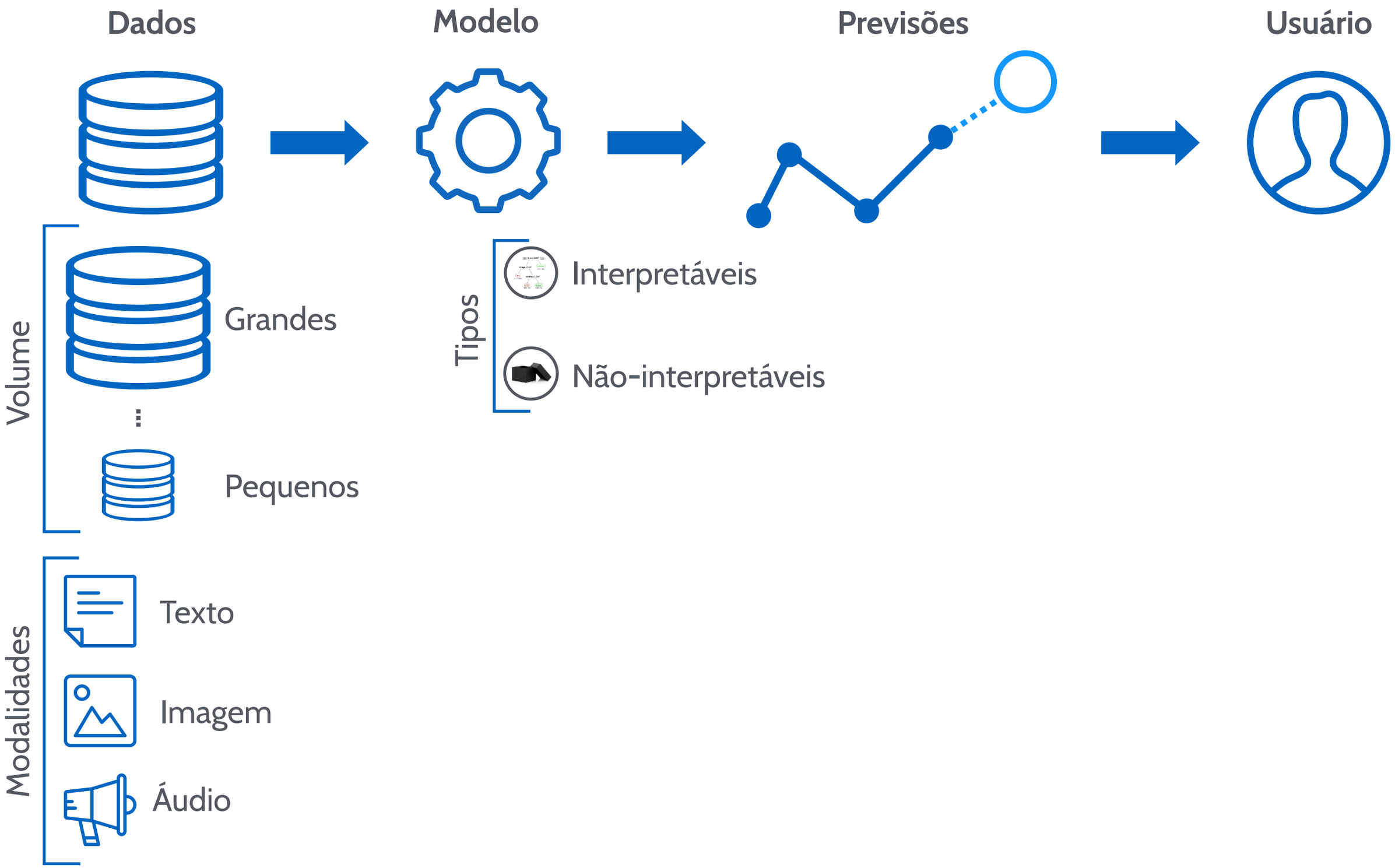


Não-interpretáveis

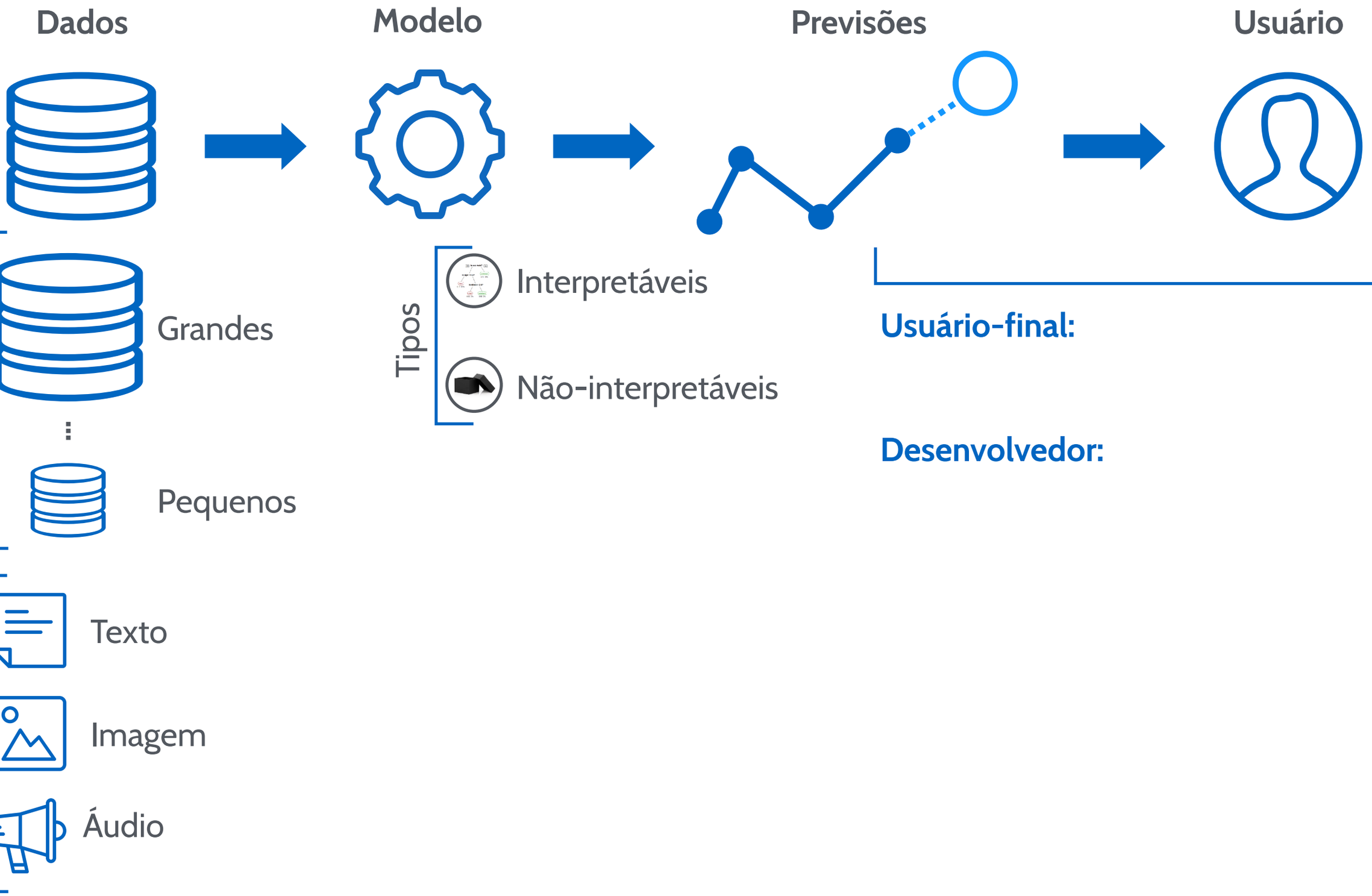
Trade-off



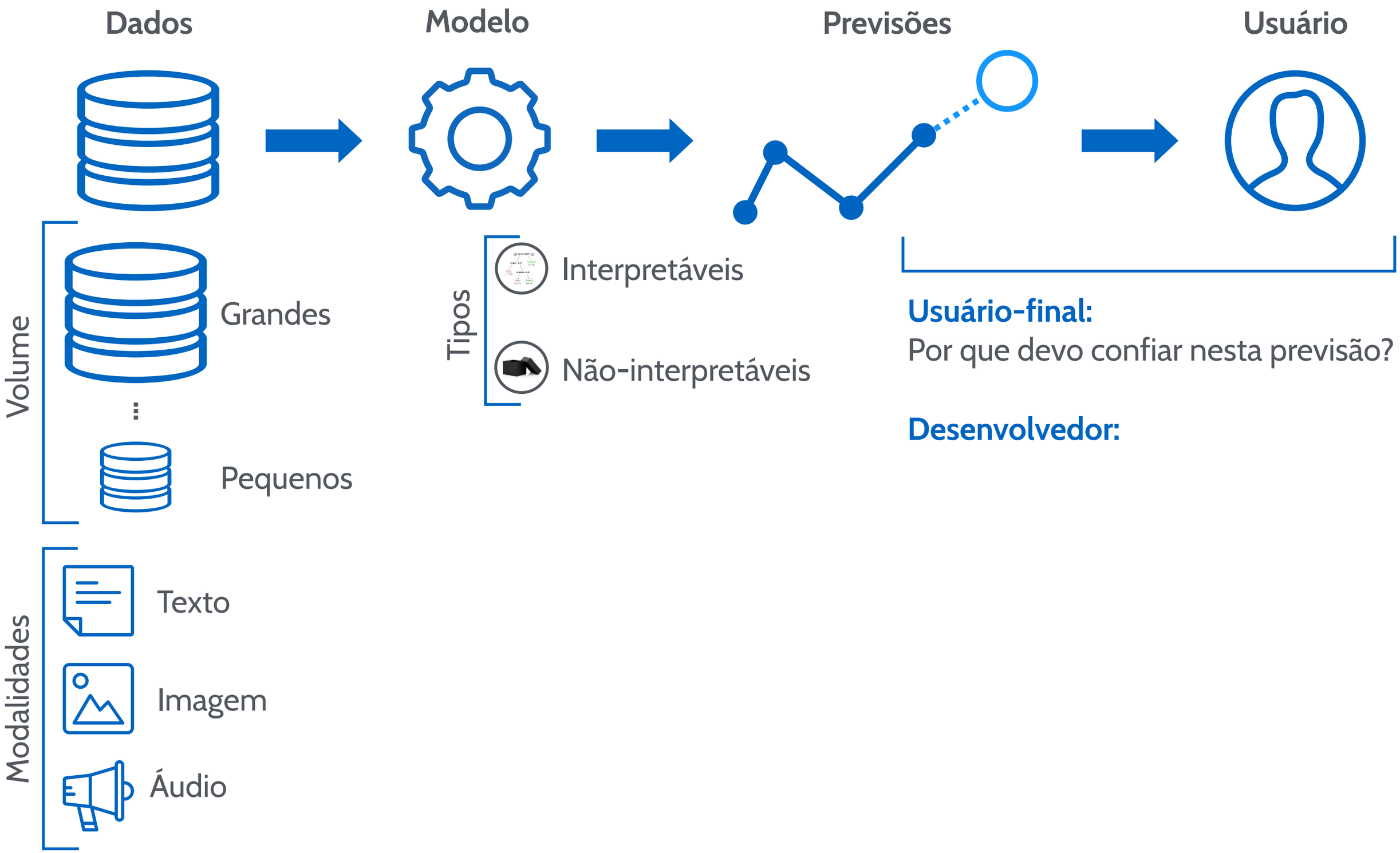
Contextualização



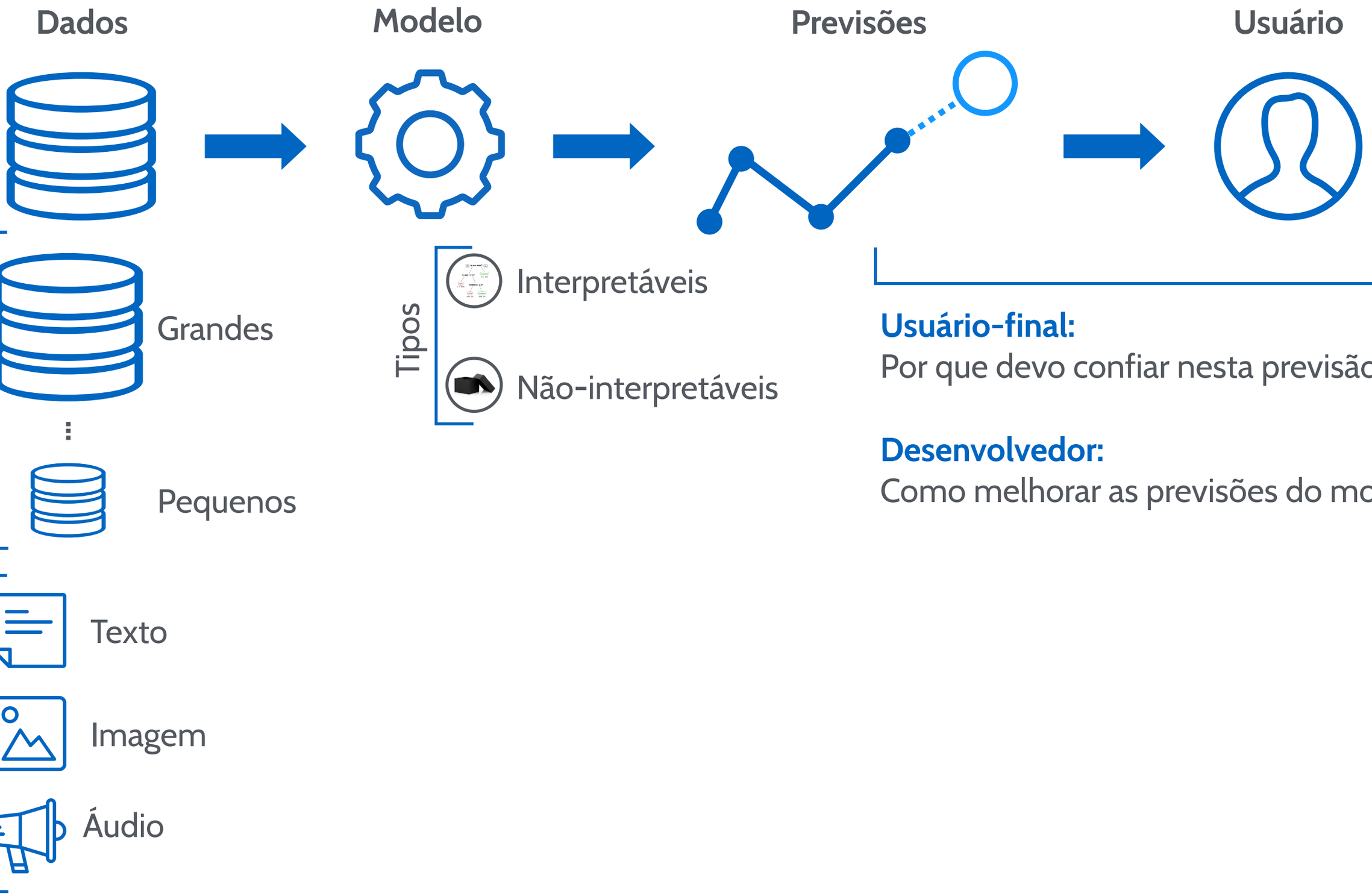
Contextualização



Contextualização



Contextualização



Usuário-final:
Por que devo confiar nesta previsão?

Desenvolvedor:
Como melhorar as previsões do modelo?

Contextualização

Dados

Modelo

Previsões

Usuário

Volume

Grandes

Pequenos

Tipos

Interpretáveis

Não-interpretáveis

Modalidades

Texto

Imagem

Áudio

Usuário-final:
Por que devo confiar nesta previsão?

Desenvolvedor:
Como melhorar as previsões do modelo?

Interpretabilidade

Contextualização

Dados

Modelo

Previsões

Usuário

Volume

Grandes

Pequenos

Tipos

Interpretáveis

Não-interpretáveis

Modalidades

Texto

Imagem

Áudio

Few-shot learning

Usuário-final:
Por que devo confiar nesta previsão?

Desenvolvedor:
Como melhorar as previsões do modelo?

Interpretabilidade

Contextualização

Dados

Modelo

Previsões

Usuário

Volume

Grandes

Pequenos

Few-shot learning

Tipos

Interpretáveis

Não-interpretáveis

Usuário-final:
Por que devo confiar nesta previsão?

Desenvolvedor:
Como melhorar as previsões do modelo?

Interpretabilidade

Modalidades

Texto

Imagem

Áudio

Aprendizado Multimodal

Objetivo

- **Utilizar** informação multimodal **para:**
  1. Melhorar a interpretabilidade das previsões
  2. *Few-shot learning*

Aprendizado multimodal

Texto

Imagem

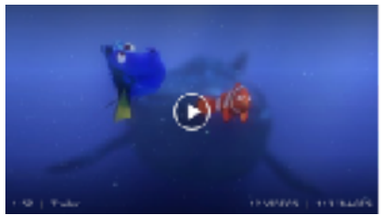
Áudio

Vídeo

Filme

After his son is captured in the Great Barrier Reef and taken to Sydney, a timid clownfish sets out on a journey to bring him home.

Trilha/
diálogos



Interpretabilidade

Definição [Miller '17]

"Capacidade de um ser humano entender uma predição"

Interpretabilidade

Confiança do usuário-final

- Por que devo confiar nestas previsões?

Melhoria do modelo por parte do desenvolvedor

- Como posso melhorar as previsões?

Interpretabilidade

- ***Right to explanation***

  Direitos dos cidadãos afetados por algoritmos de aprendizado de máquina

União européia

- *General Data Protection Regulation*
  - *"[...] the controller shall provide [...]:*
    - *the **existence of automated** decision-making and [...]*
    - *__meaningful information__ about the __logic involved__ [...]*
- 25 de maio de 2018

Estados Unidos

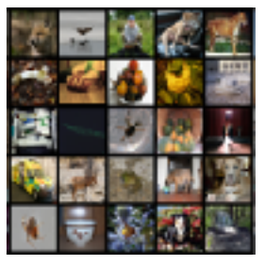
- *Equal Credit Opportunity Act*
  - *[...] creditors are required to notify applicants of action taken in certain circumstances, and such notifications must provide __specific reasons__ [...]"*

Brasil

- PLS 330/13
  - "solicitação de __revisão de decisões__ tomadas unicamente com base em tratamento automatizado de dados [...]"
- PL 5276/16
  - "o titular dos dados tem direito a solicitar __revisão__, por pessoa natural, de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses [...]"

Interpretabilidade

Hoje



Treinamento

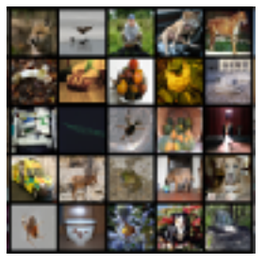
Aprendizado

Interpretabilidade

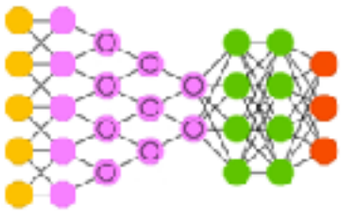
Hoje

Interpretabilidade

Hoje



Aprendizado

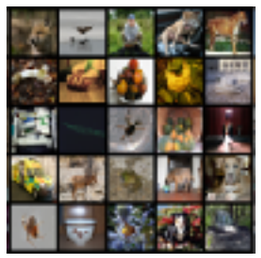


Treinamento

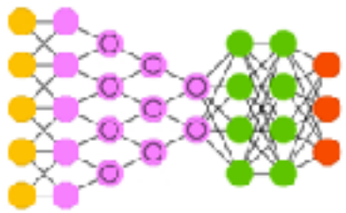
Modelo

Interpretabilidade

Hoje



Aprendizado



classe: cachorro

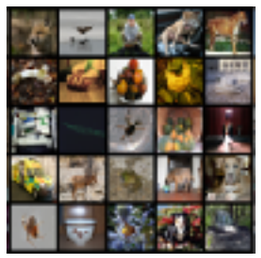
Treinamento

Modelo

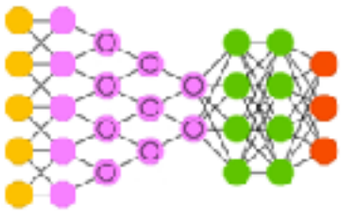
Saída

Interpretabilidade

Hoje



Aprendizado



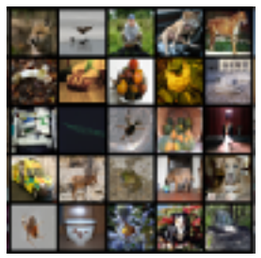
classe: cachorro

- Por que essa previsão?
- Como corrijo erros de predição?

Treinamento | Modelo | Saída | Usuário + tarefa

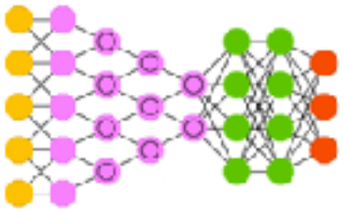
Interpretabilidade

Hoje



Treinamento

Aprendizado



Modelo

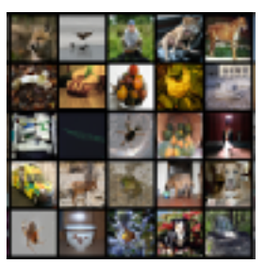
classe: cachorro

Saída

Usuário + tarefa

- Por que essa previsão?
- Como corrijo erros de predição?

Objetivo do projeto

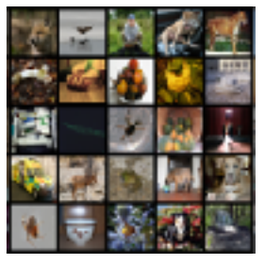


Treinamento

Novo aprendizado

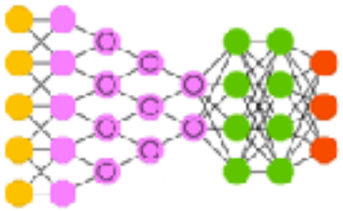
Interpretabilidade

Hoje



Treinamento

Aprendizado



Modelo

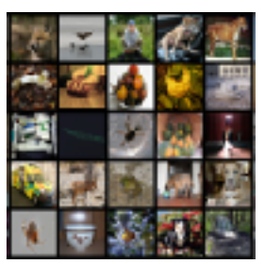
classe: cachorro

Saída

Usuário + tarefa

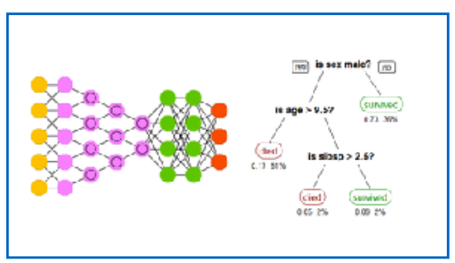
- Por que essa previsão?
- Como corrijo erros de predição?

Objetivo do projeto



Treinamento

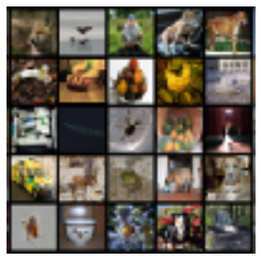
Novo aprendizado



Dois modelos

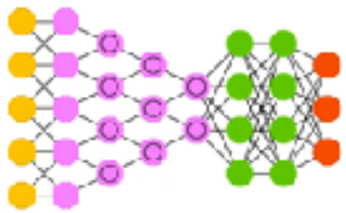
Interpretabilidade

Hoje



Treinamento

Aprendizado



Modelo

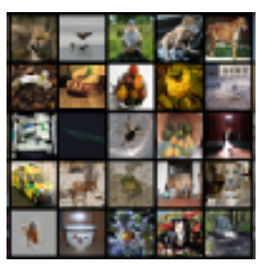
classe: cachorro

Saída

Usuário + tarefa

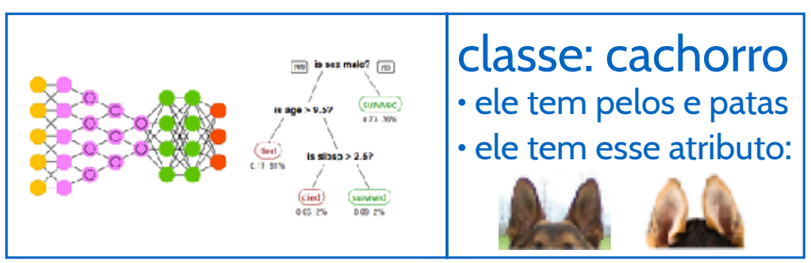
- Por que essa previsão?
- Como corrijo erros de predição?

Objetivo do projeto



Treinamento

Novo aprendizado



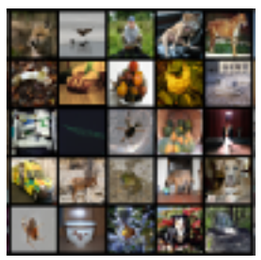
Dois modelos

classe: cachorro
- ele tem pelos e patas
- ele tem esse atributo:

Saída descritiva

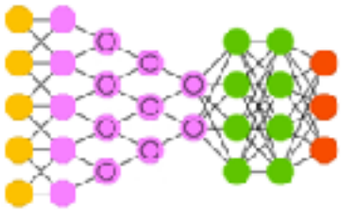
Interpretabilidade

Hoje



Treinamento

Aprendizado



Modelo

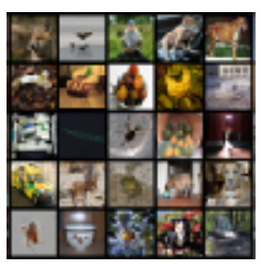
classe: cachorro

Saída

Usuário + tarefa

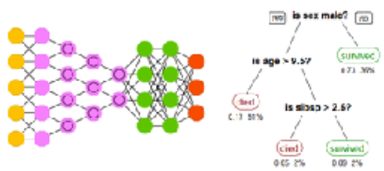
- Por que essa previsão?
- Como corrijo erros de predição?

Objetivo do projeto



Treinamento

Novo aprendizado



Dois modelos

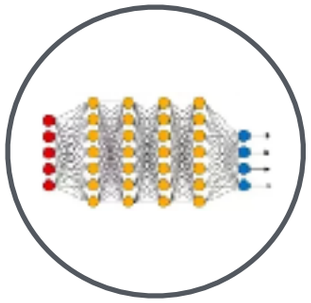
classe: cachorro
- ele tem pelos e patas
- ele tem esse atributo:

Saída descritiva

Usuário + tarefa

- Eu entendo a previsão
- Eu sei por que o modelo erra

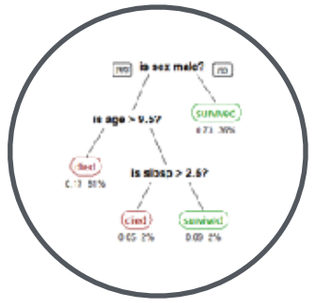
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

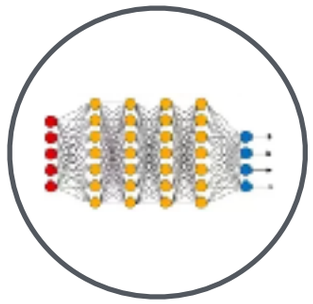
- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

Indução de modelos

- Constrói um modelo interpretável a partir de um **modelo caixa-preta**

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

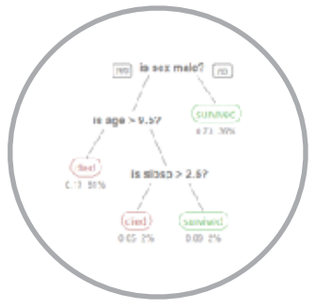
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

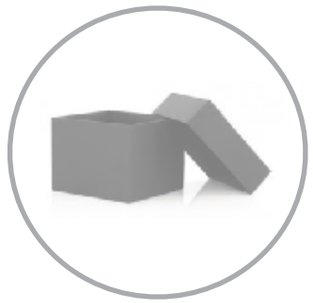
- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

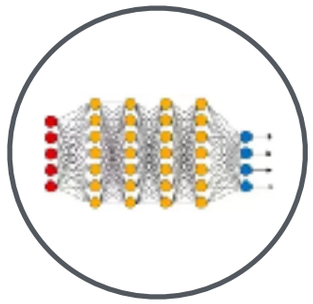


Indução de modelos

- Constrói um modelo interpretável a partir de um **modelo caixa-preta**

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

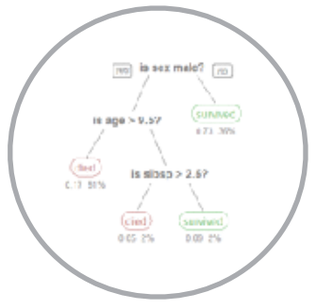
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

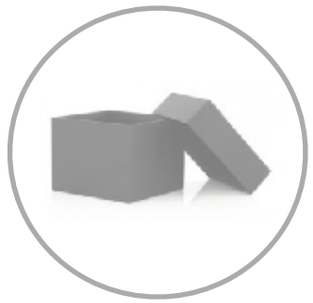
- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*



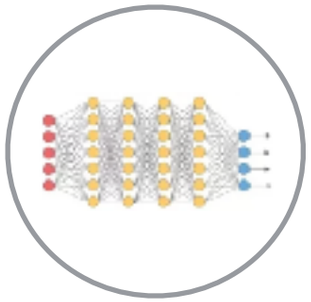
Indução de modelos

- Constrói um modelo interpretável a partir de um **modelo caixa-preta**

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

Dependentes do modelo

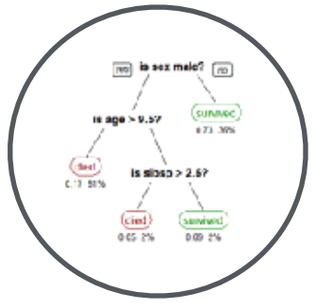
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

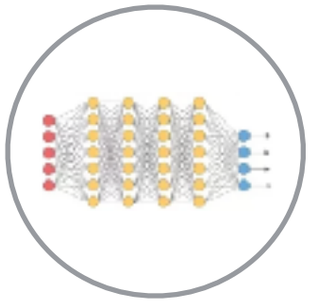
Indução de modelos

- Constrói um modelo interpretável a partir de um modelo caixa-preta

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

Dependentes do modelo

Revisão da literatura

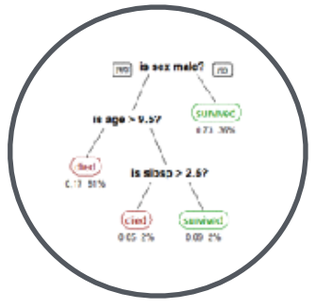


Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*

Dependentes do modelo



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

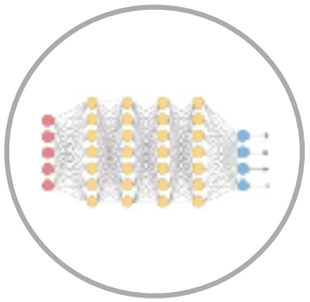
Sacrificam a acurácia

Indução de modelos

- Constrói um modelo interpretável a partir de um modelo caixa-preta

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

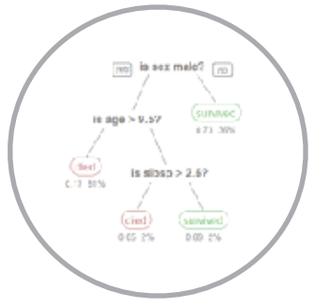
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

Indução de modelos

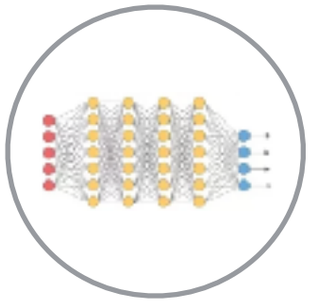
- Constrói um modelo interpretável a partir de um **modelo caixa-preta**

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

Dependentes do modelo

Sacrificam a acurácia

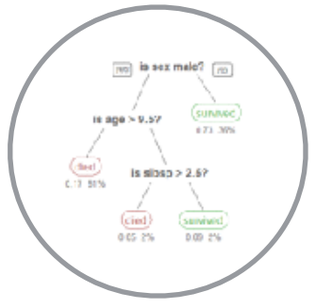
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

Indução de modelos

- Constrói um modelo interpretável a partir de um **modelo caixa-preta**

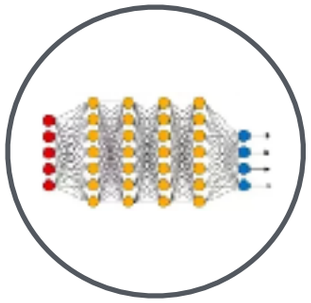
- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

Dependentes do modelo

Sacrificam a acurácia

Independente do modelo

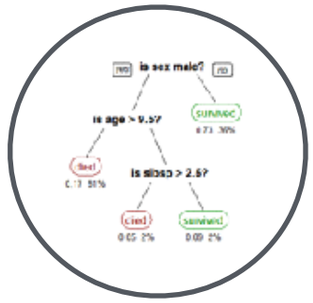
Revisão da literatura



Deep Explanations

- Modifica técnicas de aprendizado profundo para aprender características explicáveis

- *Learning Semantic Associations [Cheng et. al. '14]*
- *Learning to Generate Explanations [Hendricks et. al. '16]*



Modelos interpretáveis

- Técnicas para aprender de forma mais estruturada e interpretável

- *Bayesian Program Learning [Lake et. al. '16]*
- *Genetic Programming in Advertisement [Lacerda et. al.'06]*
- *Genetic Programming in RecSys [Oliveira et. al. '16]*

Indução de modelos

- Constrói um modelo interpretável a partir de um **modelo caixa-preta**

- *Bayesian Rule Lists [Letham et. al. '15]*
- ***Model-agnostic explanations** [Ribeiro et. al. '16]*

Dependentes do modelo

Sacrificam a acurácia

Independente do modelo

Não exploram características multimodais dos dados

Interpretação agnóstica

Previsão

f

Modelo a ser explicado

Caixa-preta

Prioriza acurácia

Foto nova

➡ cachorro

Interpretação agnóstica

Previsão

f

Modelo a ser explicado

Caixa-preta

Prioriza acurácia

Foto nova

cachorro

Explicação

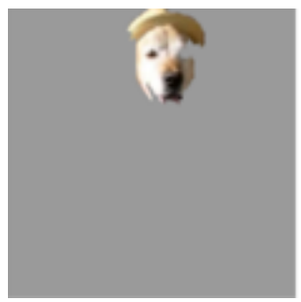
g

Modelo explicador

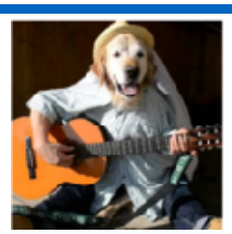
Modelos interpretáveis

Prioriza interpretabilidade

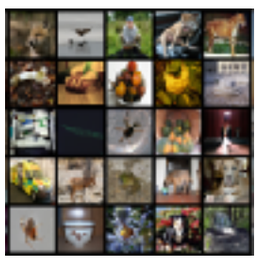
Foto nova



Indução agnóstica de modelos

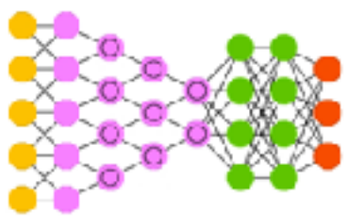


Teste



Treinamento

Aprendizado

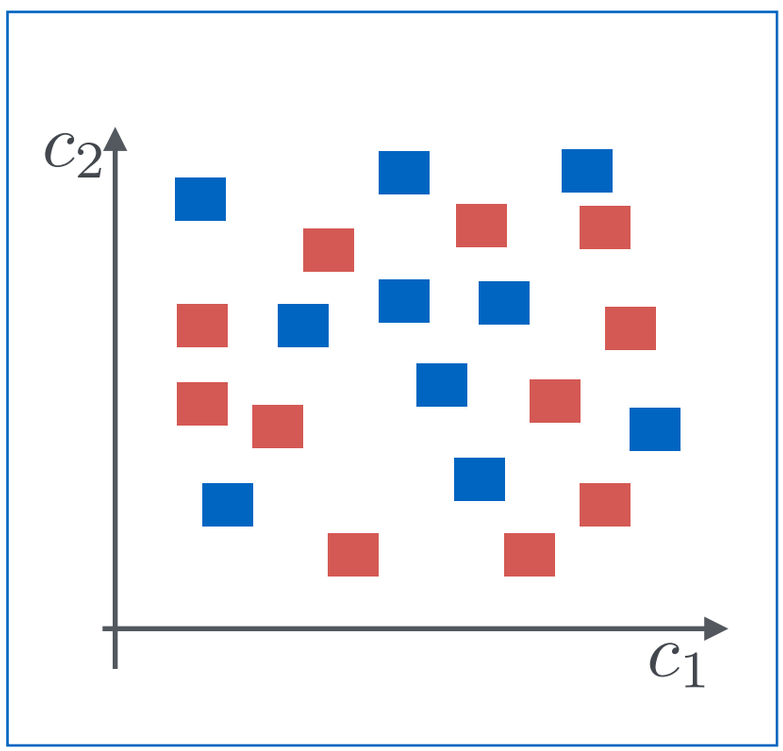


Modelo

classe: cachorro

Saída

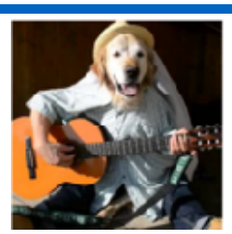
Usuário + tarefa



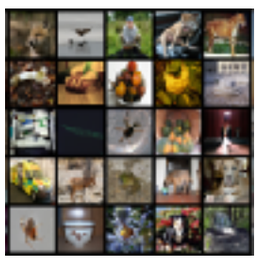
cachorro

não-cachorro

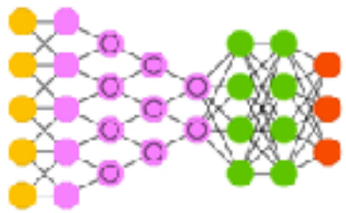
Indução agnóstica de modelos



Teste

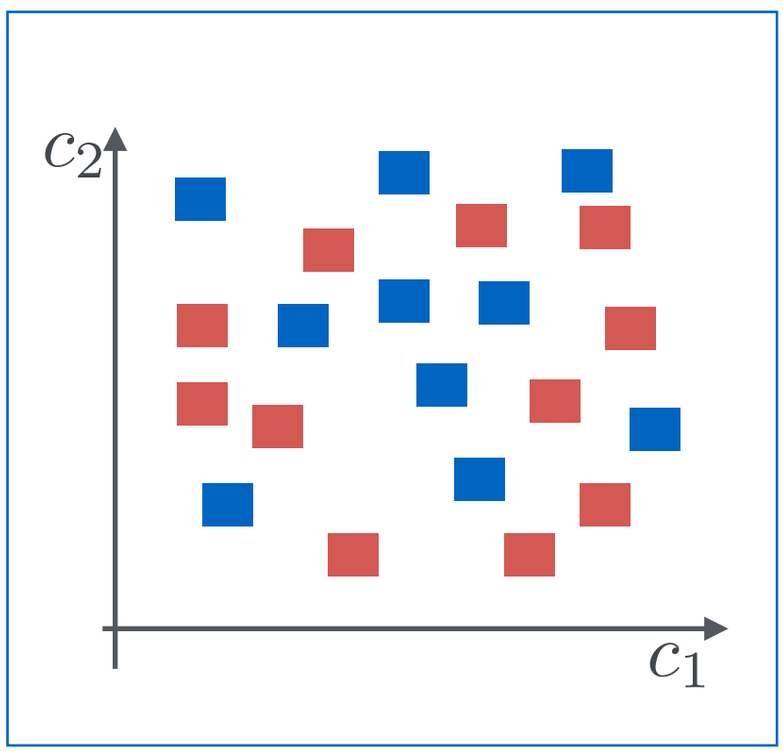


Aprendizado

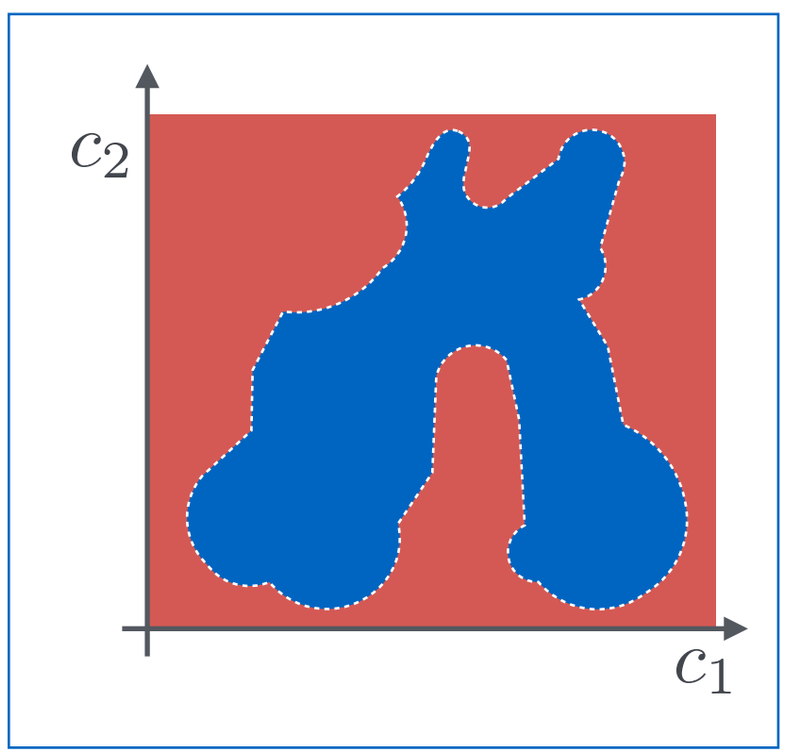


classe: cachorro

Treinamento



Modelo



Saída

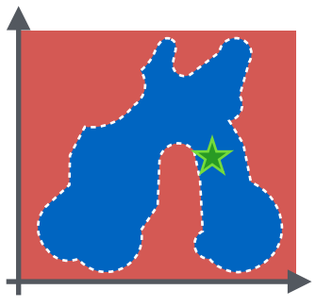
Usuário + tarefa

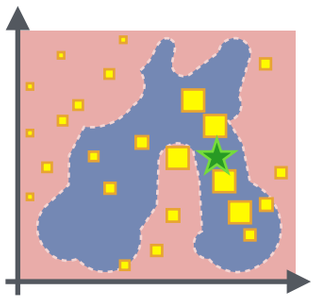
cachorro

não-cachorro

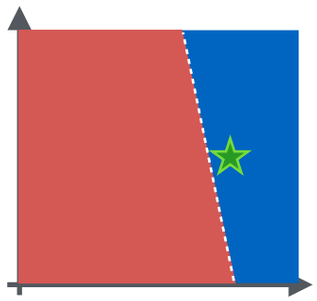
Algoritmo de indução [Ribeiro et al. '16]

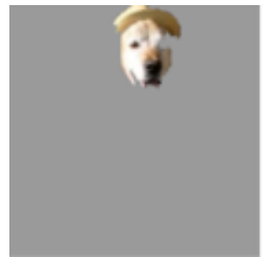
1 Instância → 2 Vizinhança → 3 Instâncias interpretáveis → 4 Modelo explicador → 5 Explicação





Instância modificada	P(cachorro)
	0,92
	0,01
	0,34





Algoritmo de indução [Ribeiro et al. '16]

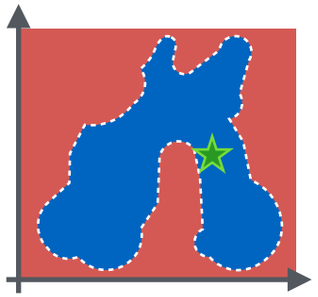
1 Instância

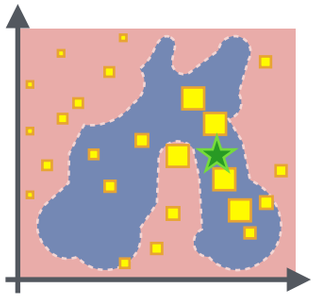
2 Vizinhança

3 Instâncias interpretáveis

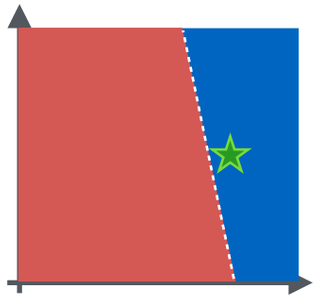
4 Modelo explicador

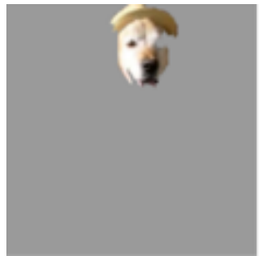
5 Explicação





Instância modificada	P(cachorro)
	0,92
	0,01
	0,34





Indução de um modelo explicador

1

Instância de interesse

Previsão
f

Modelo a ser explicado

Caixa-preta

Prioriza acurácia

Características não-interpretáveis

Foto nova

cachorro

1

Instância de interesse

Previsão
f

Modelo a ser explicado

Caixa-preta

Prioriza acurácia

Características não-interpretáveis

Foto nova

➡ cachorro

Instâncias

Características não-interpretáveis (x)			Classe
C_1	...	C_d	Y
	...		
	...		

- Tensor RGB

1

Instância de interesse

Previsão
f

Modelo a ser explicado

Caixa-preta

Prioriza acurácia

Características não-interpretáveis

Foto nova

➡ cachorro

Instâncias

Características não-interpretáveis (x)			Classe
C_1	...	C_d	Y
	...		
	...		

- Tensor RGB

$$Y = f(x)$$

2 Vizinhança de interesse

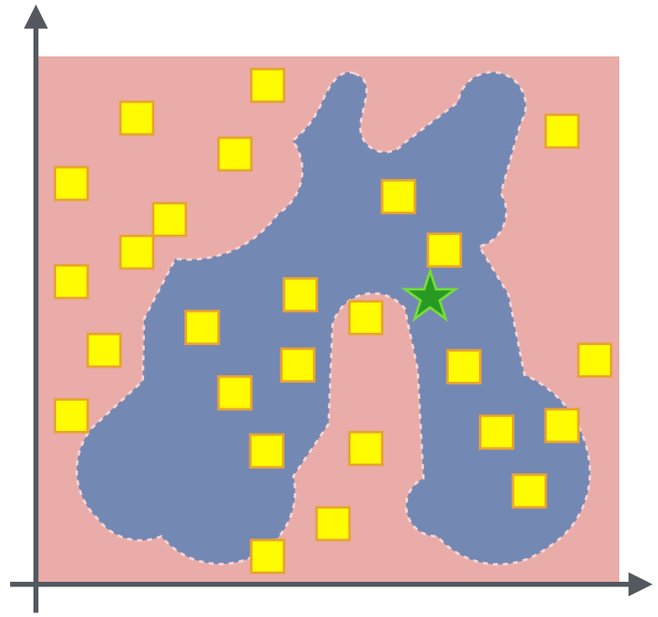
- **Amostre pontos**

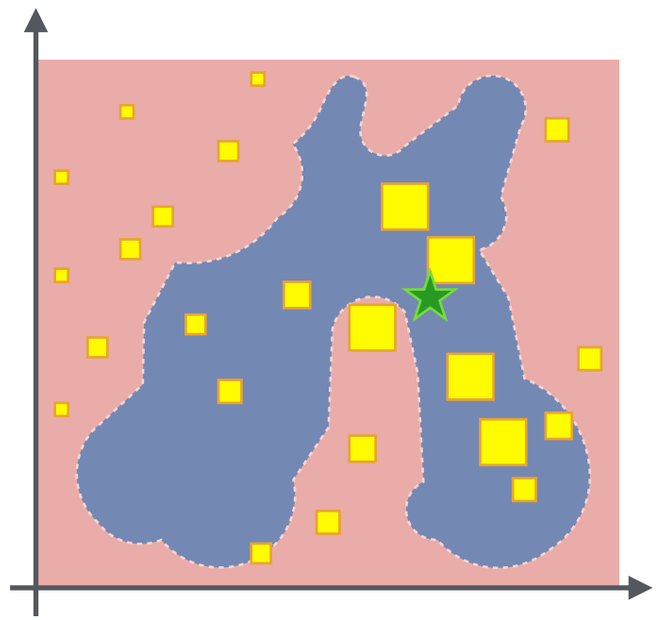
  Distribuição normal (centro de massa)

- **Pondere o peso das instâncias**

  De acordo com uma métrica de similaridade

  - Texto: cosseno
  - Imagem: L2





3 Representação interpretável

- **Nova base de dados**

  Características interpretáveis

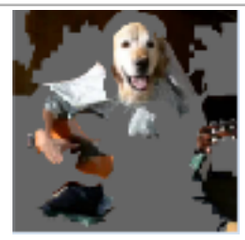
· *Superpixels*

- **Ex: imagens**

  Elimina partes da imagem

- **Ex: textos**

  Elimina palavras do texto

Instância modificada	P(cachorro)
	0,92
	0,25
	0,54

3 Representação interpretável

- **Nova base de dados**

  Características interpretáveis

Características não-interpretáveis (x)

Instâncias	C_1	...	C_d
		...	
		...	

- Tensor RGB
- *Word embeddings*

3 Representação interpretável

- **Nova base de dados**

  Características interpretáveis

Instâncias

Características não-interpretáveis (x)		
C_1	...	C_d
	...	
	...	

- Tensor RGB
- *Word embeddings*

f

Características interpretáveis (x')		
$C_{1'}$	...	$C_{d'}$
	...	
	...	

- Superpixels
- Palavras

g

4 Treine modelo explicador

- **Modelo interpretável (g)**

  Modelo linear: regressão logística

- **Seleção de características (K)**

  Escolha do número de características no modelo explicador

Características interpretáveis (z')

Classe

Instâncias

C_1	...	C_K	f(z)
	...		
	...		

Indução

- Superpixels
- Palavras

g

4 Treine modelo explicador

- **Modelo interpretável (g)**

  Modelo linear: regressão logística

- **Seleção de características (K)**

  Escolha do número de características no modelo explicador

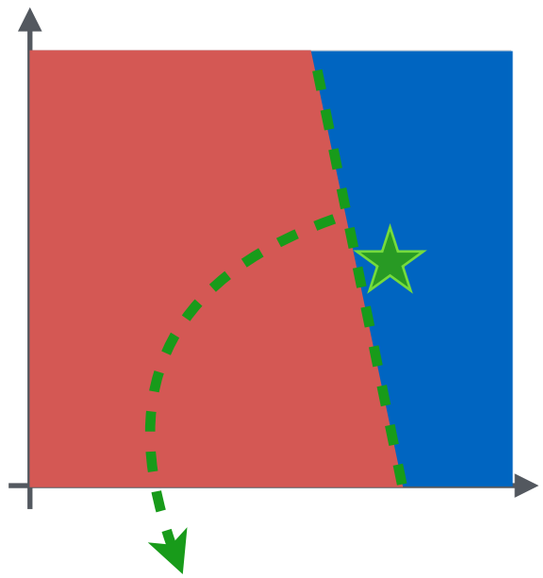
Características interpretáveis (z')

Classe

Instâncias

C_1	...	C_K	f(z)
	...		
	...		

- Superpixels
- Palavras



$$-10\,c_1 - 2\,c_2 + 5$$

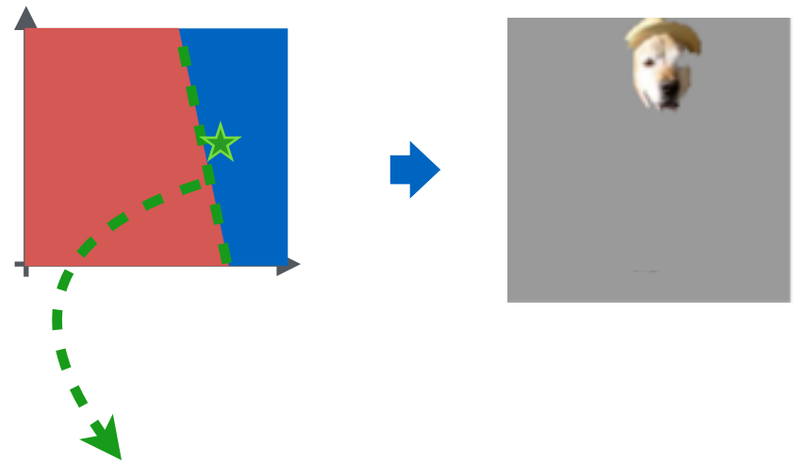
5

Explique a previsão

Explicação
g

Modelo explicador

Regressão logística
Super-pixels



$$-10\,c_1 - 2\,c_2 + 5$$

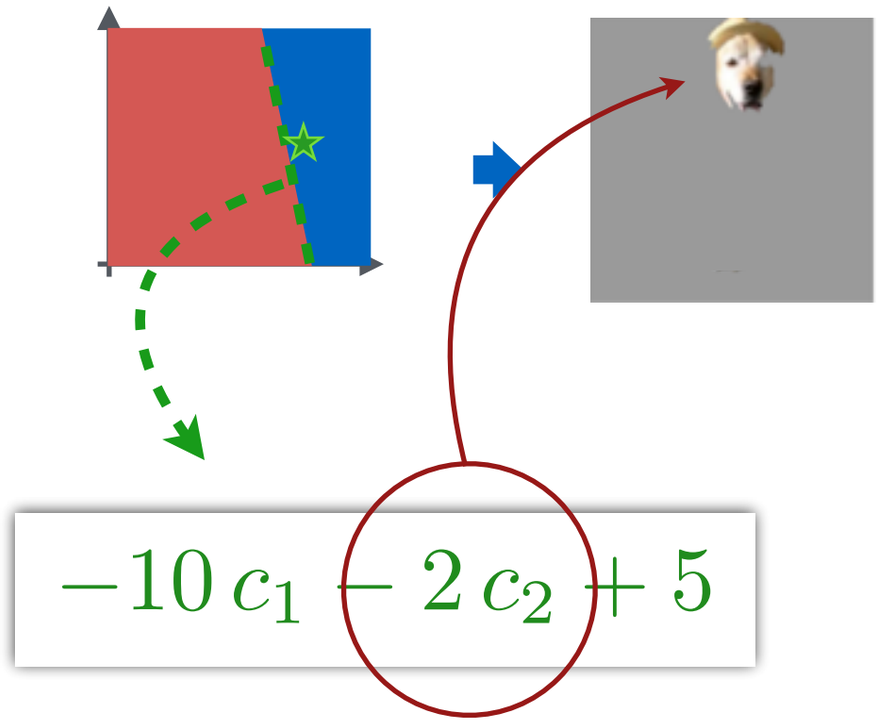
5

Explique a previsão

Explicação g

Modelo explicador

Regressão logística
Super-pixels



Indução agnóstica de modelos interpretáveis

- **Formalmente**

  Dada a instância x a ser explicada, a explicação $\xi(x)$ é dada por:

$$\xi(x) = \underset{g \in G}{\arg\min} \quad \overbrace{L(f, g, \pi_x)}^{\text{erro ponderado}} \quad + \quad \overbrace{\Omega(g)}^{\text{complexidade}}$$

f : modelo a ser explicado
g : modelo "explicador"
π_x : vizinhança de x

Indução agnóstica de modelos interpretáveis

- **Formalmente**

  Dada a instância x a ser explicada, a explicação $\xi(x)$ é dada por:

$$\xi(x) = \underset{g \in G}{\arg\min} \quad \overbrace{L(f, g, \pi_x)}^{\text{erro ponderado}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

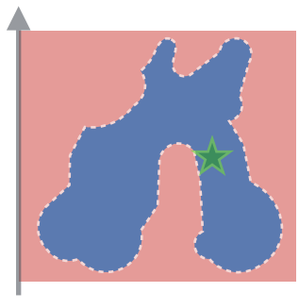
$$= \underset{g \in G}{\arg\min} \sum_{z, z' \in Z} \overbrace{\pi_x(z)}^{\text{peso}} \overbrace{(f(z) - g(z'))^2}^{\text{erro}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

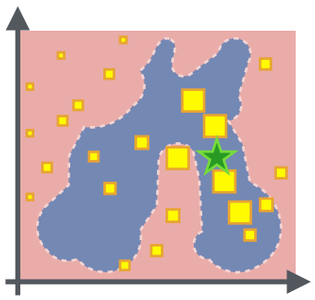
f : modelo a ser explicado
g : modelo "explicador"
π_x : vizinhança de x

Limitações

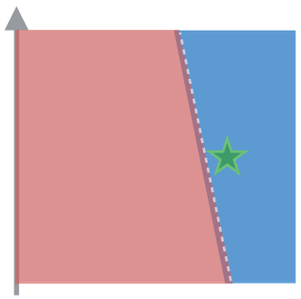
Limitações

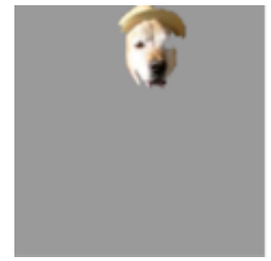
1 Instância → 2 Vizinhança → 3 Instâncias interpretáveis → 4 Modelo explicador → 5 Explicação





Instância modificada	P(cachorro)
	0,92
	0,01
	0,34





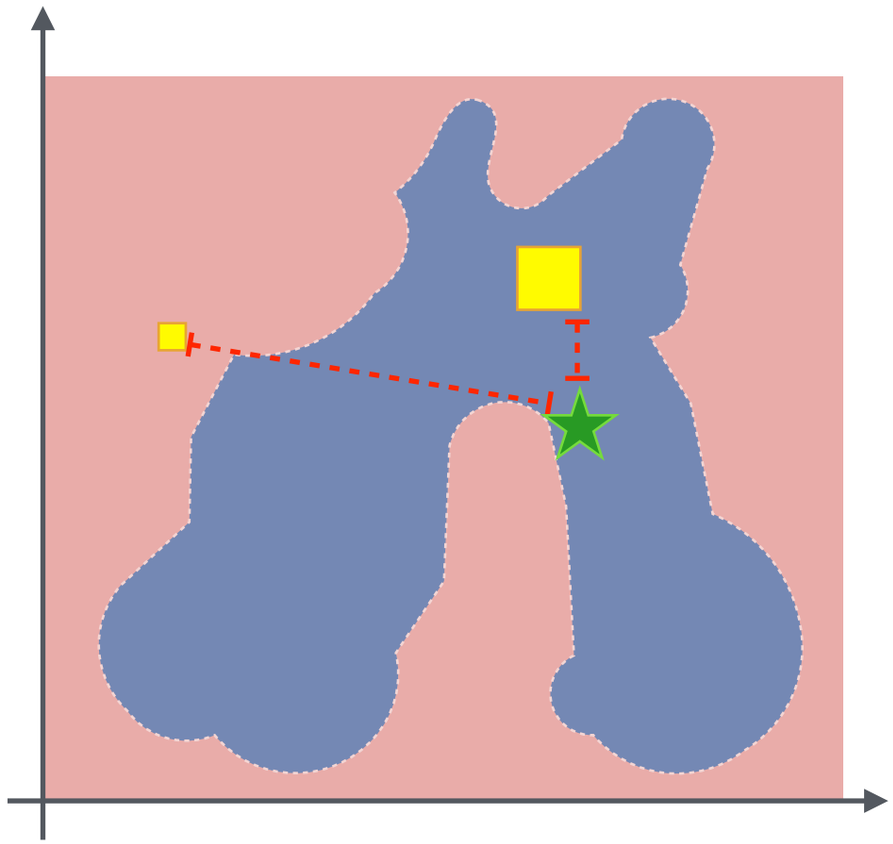
- **Limitação 1: encontrar vizinhança relevante**

  Métricas de distância dependente da modalidade

Limitação: definir vizinhança

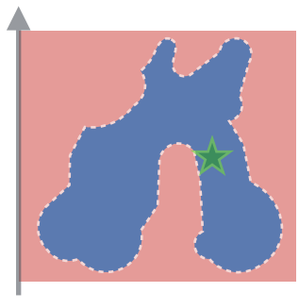
- **Ponderação de acordo com a similaridade**

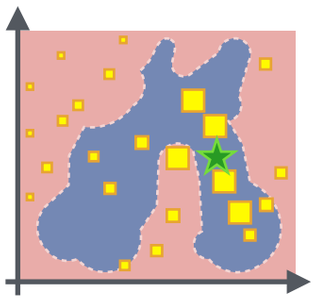
  Imagem: métrica L2



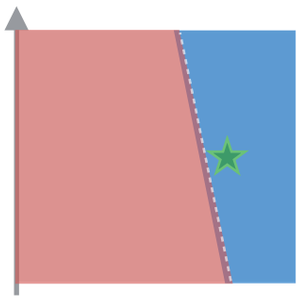
Limitações

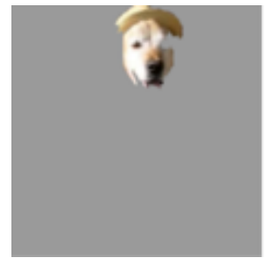
1 Instância → 2 Vizinhança → 3 Instâncias interpretáveis → 4 Modelo explicador → 5 Explicação





Instância modificada	P(cachorro)
	0,92
	0,01
	0,34





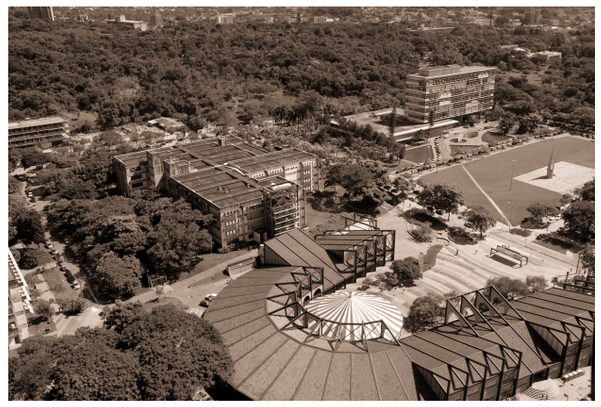
- **Limitação 1: encontrar vizinhança relevante**

  Métricas de distância dependente da modalidade

- **Limitação 2: explicar a previsão**
  - Explicação de **imagens**: **superpixels**

Limitação: explicação da predição

- **Classes:** sepia não-sepia

Limitação: explicação da predição

- **Classes:** sepia não-sepia

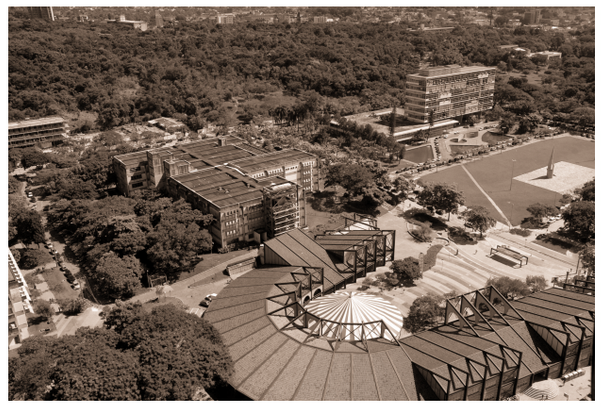



Foto nova

f ➡ Previsão

g ➡ Explicação

Interpretação agnóstica

Previsão

f

Modelo a ser explicado

Prioriza acurácia

Foto nova

sepia

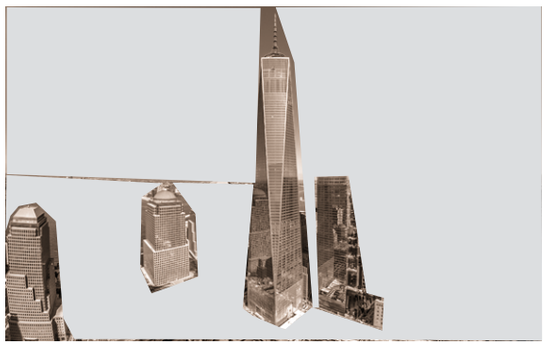
Explicação

g

Modelo explicador

Prioriza interpretabilidade

Foto nova



Interpretação agnóstica

Previsão
f

Modelo a ser explicado

Prioriza acurácia

Foto nova

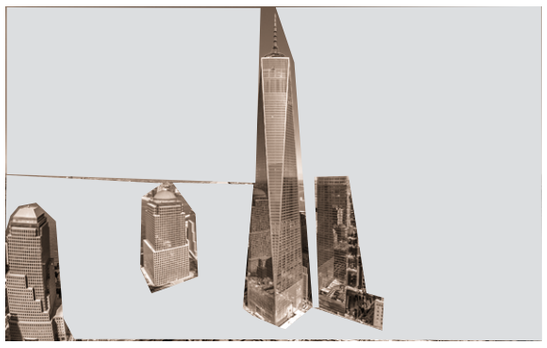
sepia

Explicação
g

Modelo explicador

Prioriza interpretabilidade

Foto nova



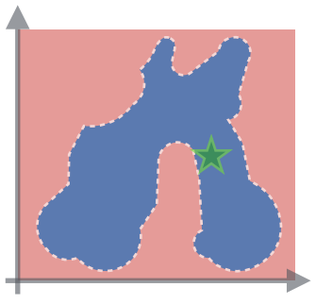
- ***Superpixels* explicam partes da imagem**

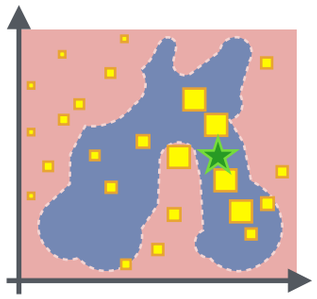
  Não descrevem características globais como tom sepia

Abordagem proposta

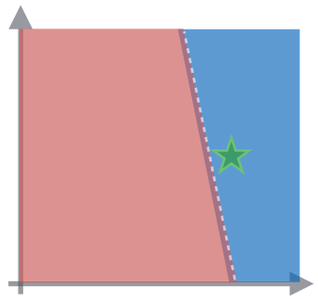
Indução multimodal

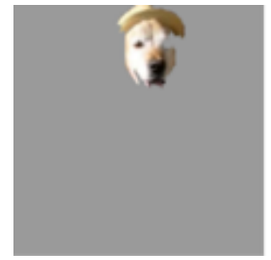
1. Instância
2. Vizinhança
3. Instâncias interpretáveis
4. Modelo explicador
5. Explicação

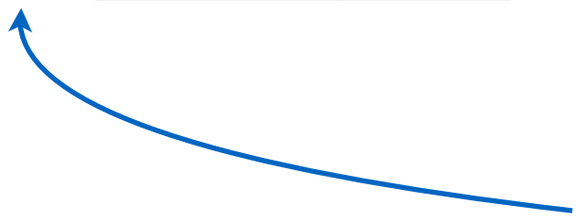




Instância modificada	P(cachorro)
	0,92
	0,01
	0,34







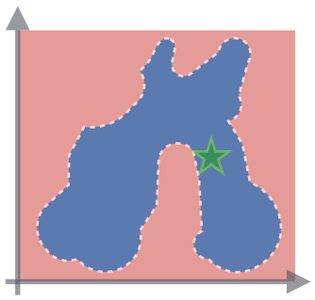
<texto>

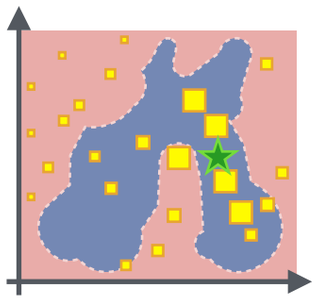
cachorro
raça
cocker
spaniel

mulher
ruiva
sorrindo

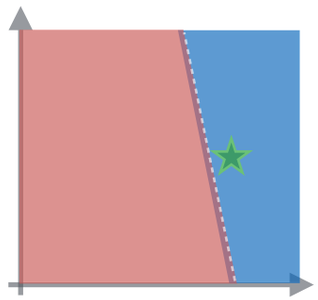
Indução multimodal

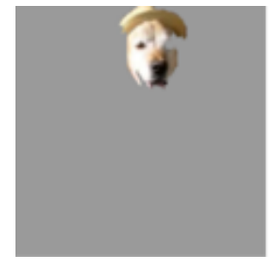
1 Instância → 2 Vizinhança → 3 Instâncias interpretáveis → 4 Modelo explicador → 5 Explicação





Instância modificada	P(cachorro)
	0,92
	0,01
	0,34

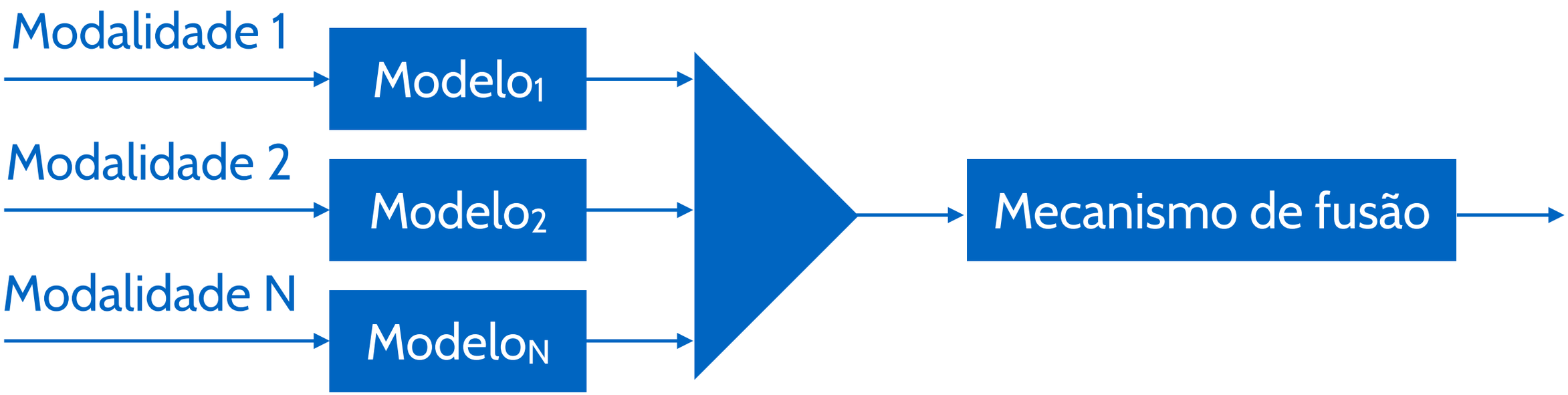




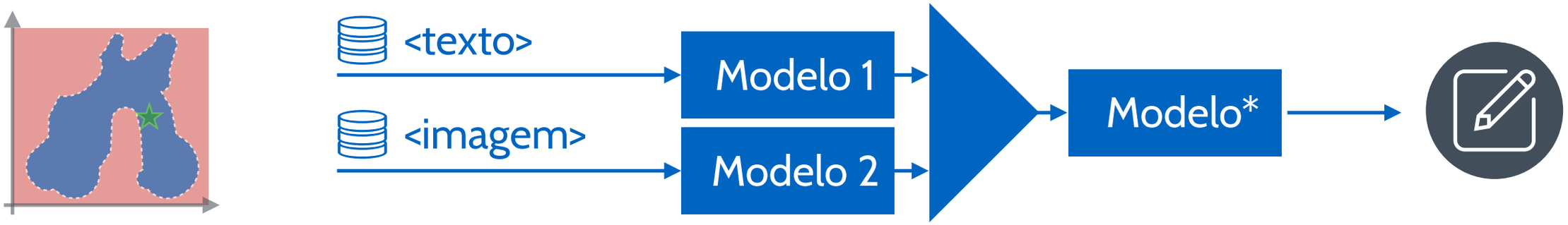
- **Representação multimodal:**
  - Fusão tardia
  - Fusão antecipada

Fusão [Baltrusaitis et al. '18]

- **Fusão tardia (*late fusion)***
  - Treinar um modelo para cada modalidade
  - Múltiplas etapas de treinamento (em paralelo)
  - Mecanismo de fusão: votação, soma ou um método de aprendizado



Indução multimodal: *fusão tardia*

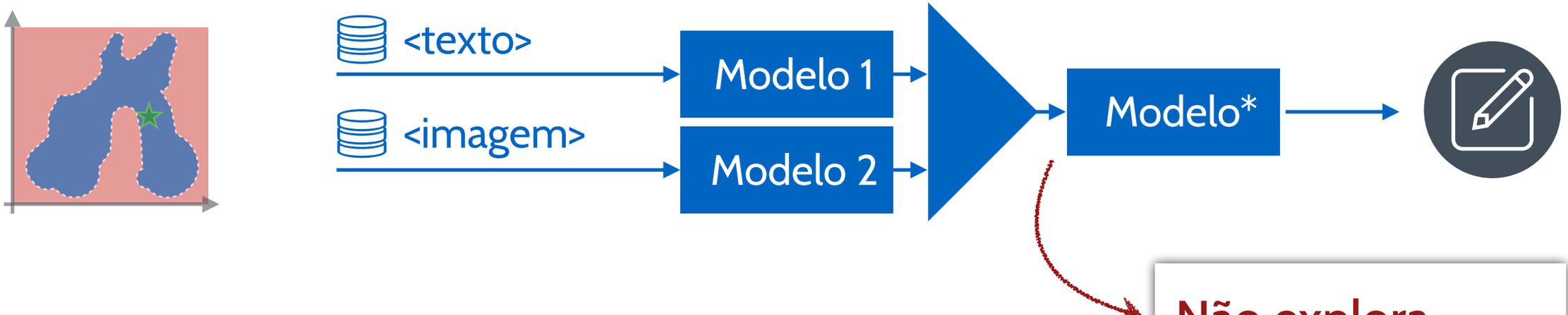


- **Métrica de vizinhança**
  - Texto: cosseno
  - Imagem/vídeo: métrica L2
- **Modelo explicador**

  Modelo com mais acertos dentre os modelos de vizinhança

Indução multimodal: *fusão tardia*

1. Instância
2. Vizinhança
3. Instâncias interpretáveis
4. Modelo explicador
5. Explicação

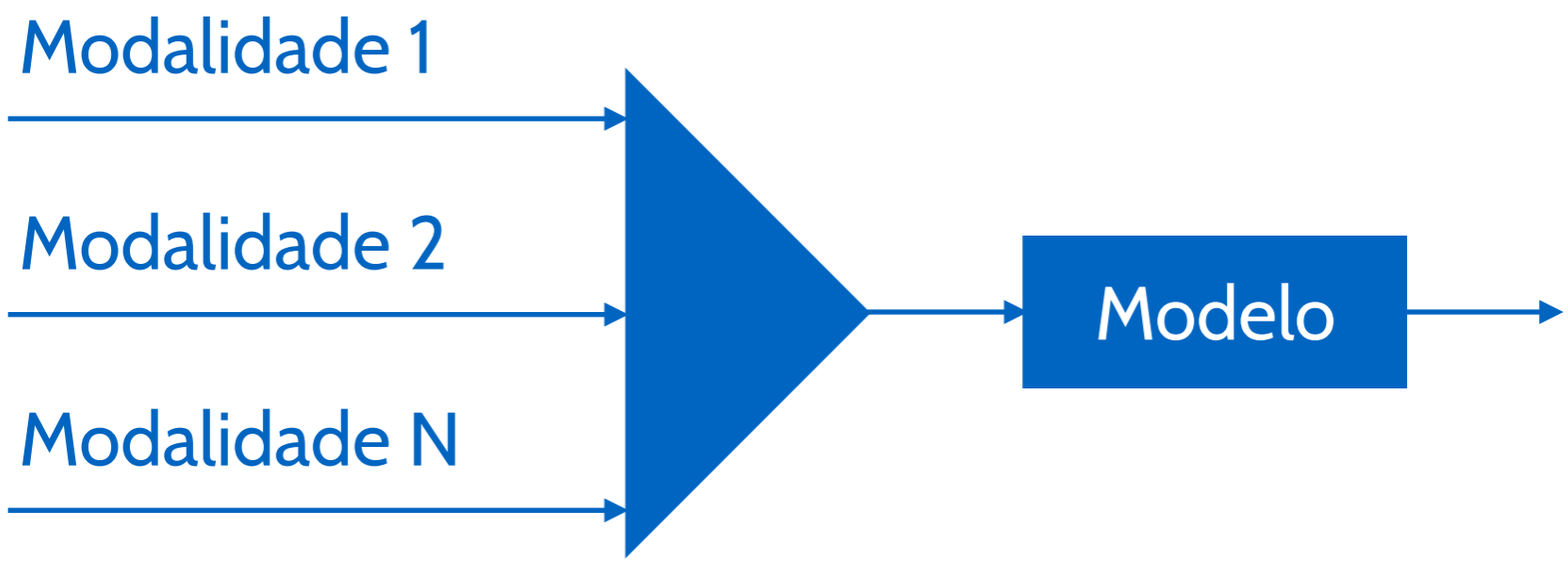


- **Métrica de vizinhança**
  - Texto: cosseno
  - Imagem/vídeo: métrica L2
- **Modelo explicador**

  Modelo com mais acertos dentre os modelos de vizinhança

Fusão [Baltrusaitis et al. ’18]

- **Fusão antecipada (*early fusion)***
  - Fácil implementação (concatenação)
  - Explora dependências entre as características



Indução multimodal: fusão antecipada

$< \texttt{visual}_1 + \texttt{textual}_1 >$

cachorro
raça
cocker
spaniel

?

$< \texttt{visual}_2 + \texttt{textual}_2 >$

mulher
ruiva
sorrindo

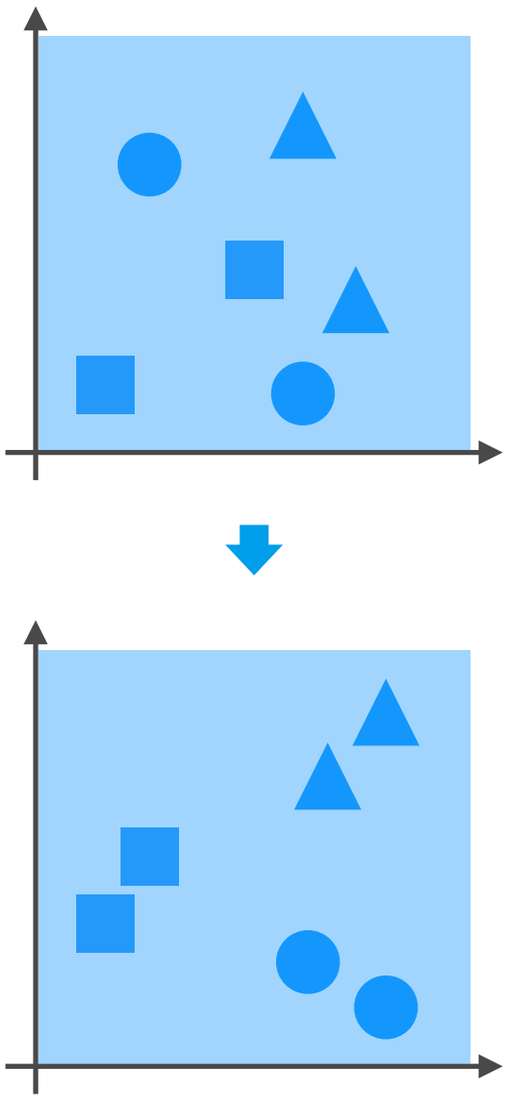
- **Métrica de vizinhança**
  - Como comparar vetores de várias modalidades?
  - $D\left(< \texttt{visual}_1 + \texttt{textual}_1 >, < \texttt{visual}_2 + \texttt{textual}_2 >\right)$

- ***Distance Metric Learning*** **[Xing et. al. '03]**

Distance metric learning [Xing et al. '03]

Instâncias:

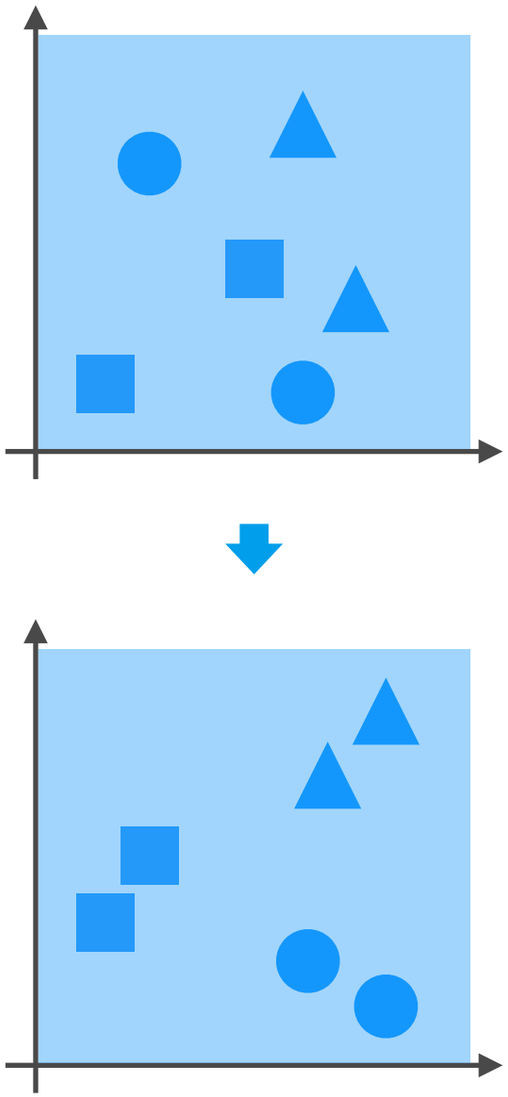
- similares ⟶ próximas
- dissimilares ⟶ afastadas

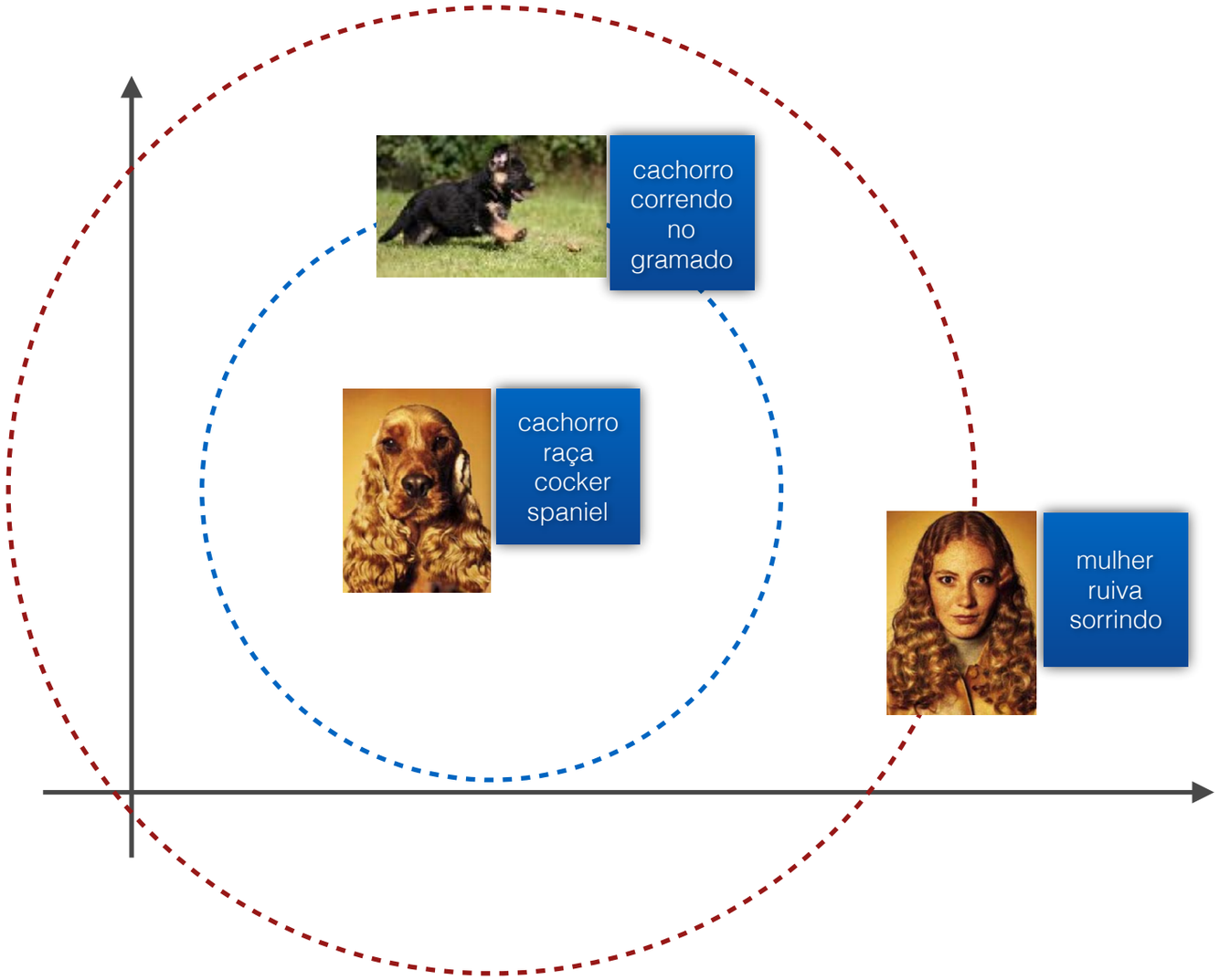


Distance metric learning [Xing et al. '03]

Instâncias:

- similares ⟶ próximas
- dissimilares ⟶ afastadas





Distance metric learning [Xing et al. '03]

- **Sejam:** $\begin{cases} S : (x_i, x_j) \in S & \text{se } x_i \text{ e } x_j \text{ são similares} \\ D : (x_i, x_j) \in D & \text{se } x_i \text{ e } x_j \text{ são dissimilares} \end{cases}$

Distance metric learning [Xing et al. '03]

- **Sejam:** $\begin{cases} S : (x_i, x_j) \in S & \text{se } x_i \text{ e } x_j \text{ são similares} \\ D : (x_i, x_j) \in D & \text{se } x_i \text{ e } x_j \text{ são dissimilares} \end{cases}$

- **Aprendizado de uma métrica de distância da forma:**

$$M(x, z) = M_A(x, z) = \|x - z\|_A = \sqrt{(x - z)A(x - z)}$$

  - Se $A = \mathcal{I}$: distância L2 (Euclidiana)
  - A parametriza uma família de métricas sobre $\mathbb{R}^n$, onde n é o número de características das instâncias

Distance metric learning [Xing et al. '03]

- **Sejam:** $\begin{cases} S : (x_i, x_j) \in S & \text{se } x_i \text{ e } x_j \text{ são similares} \\ D : (x_i, x_j) \in D & \text{se } x_i \text{ e } x_j \text{ são dissimilares} \end{cases}$

- **Aprendizado de uma métrica de distância da forma:**

$$M(x, z) = M_A(x, z) = \|x - z\|_A = \sqrt{(x - z)A(x - z)}$$

  - Se $A = \mathcal{I}$: distância L2 (Euclidiana)
  - A parametriza uma família de métricas sobre $\mathbb{R}^n$, onde n é o número de características das instâncias

- **Função de aprendizado:**

$$\tilde{A} = \arg\min_A \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_A + \sum_{(x_i, x_j) \in D} \max(0, 1 - \|x_i - x_j\|_A)$$

Indução multimodal + *metric learning*

- **Indução de modelos**

$$\xi(x) = \arg\min_{g \in G} \; L(f, g, \pi_x) \; + \; \Omega(g)$$

$$= \arg\min_{g \in G} \sum_{z,z' \in Z} \overbrace{\pi_x(z)}^{\text{peso}} \overbrace{(f(z) - g(z'))^2}^{\text{erro}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

Indução multimodal + *metric learning*

- **Indução de modelos**

$$\xi(x) = \arg\min_{g \in G} \; L(f, g, \pi_x) \; + \; \Omega(g)$$

$$= \arg\min_{g \in G} \sum_{z,z' \in Z} \overbrace{\pi_x(z)}^{\text{peso}} \overbrace{(f(z) - g(z'))^2}^{\text{erro}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

- **Indução de modelos multimodal + *metric learning***

$$\xi(x) = \arg\min_{g \in G} \sum_{z,z' \in Z} \overbrace{M_{\tilde{A}}(x, z)}^{\text{peso}} \overbrace{(f(z) - g(z'))^2}^{\text{erro}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

$$\tilde{A} = \arg\min_{A} \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_A + \sum_{(x_i, x_j) \in D} \max(0, 1 - \|x_i - x_j\|_A)$$

Indução multimodal + *metric learning*

- **Indução de modelos**

$$\xi(x) = \arg\min_{g \in G} \; L(f, g, \pi_x) \; + \; \Omega(g)$$

$$= \arg\min_{g \in G} \sum_{z, z' \in Z} \overbrace{\pi_x(z)}^{\text{peso}} \overbrace{(f(z) - g(z'))^2}^{\text{erro}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

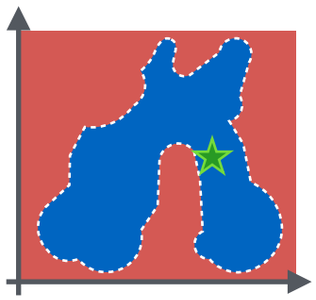
- **Indução de modelos multimodal + *metric learning***

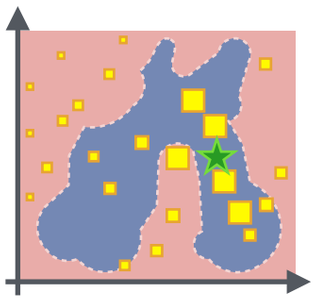
$$\xi(x) = \arg\min_{g \in G} \sum_{z, z' \in Z} \overbrace{M_{\tilde{A}}(x, z)}^{\text{peso}} \overbrace{(f(z) - g(z'))^2}^{\text{erro}} + \overbrace{\Omega(g)}^{\text{complexidade}}$$

$$\tilde{A} = \arg\min_{A} \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_A + \sum_{(x_i, x_j) \in D} \max(0, 1 - \|x_i - x_j\|_A)$$

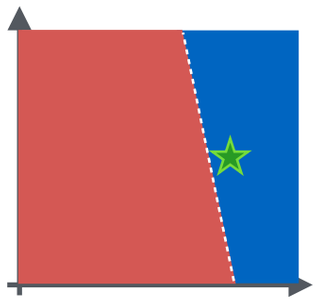
Limitações

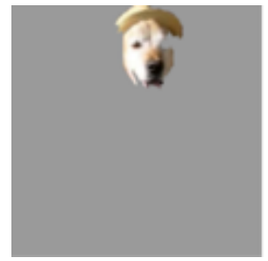
1 Instância → 2 Vizinhança → 3 Instâncias interpretáveis → 4 Modelo explicador → 5 Explicação





Instância modificada	P(cachorro)
	0,92
	0,01
	0,34





- **Limitação 1: encontrar vizinhança relevante**

  Métricas de distância dependente da modalidade

- **Limitação 2: explicar a previsão**
  - Explicação de **imagens**: **superpixels**

Explicação por exemplos [Kim et al. '16]

- **Seleção de exemplos**
  - similares
  - dissimilares

Explicação por exemplos [Kim et al. '16]

- **Seleção de exemplos**
  - similares ➡ entre os que o usuário gostou
  - dissimilares ➡ entre os que o usuário não gostou

Usuário-alvo

➡ Sistema de Recomendação ➡ Recomendação

Explicação por exemplos [Kim et al. '16]

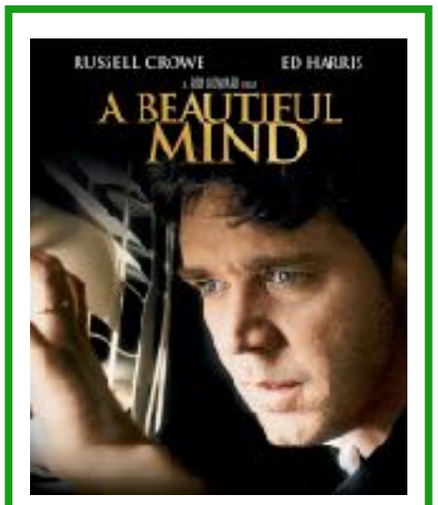
- **Seleção de exemplos**
  - similares
  - dissimilares

Usuário-alvo

Sistema de Recomendação

Recomendação

Por que você gostou de:



Por que você não gostou de:

Explicação por exemplos [Kim et al. '16]

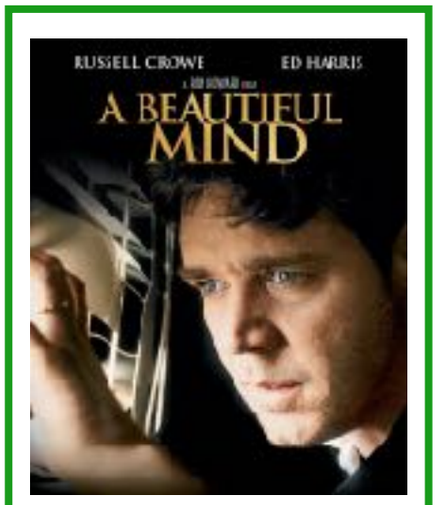
- **Seleção de exemplos**
  - similares
  - dissimilares

Usuário-alvo

Sistema de Recomendação

Recomendação

Por que você gostou de:



Por que você não gostou de:

Explicação por exemplos [Kim et al. '16]

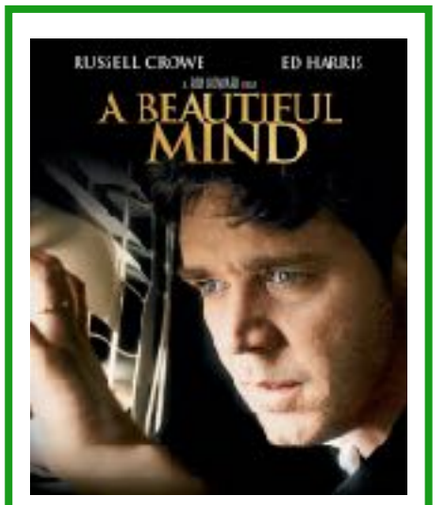
- **Seleção de exemplos**
  - similares
  - dissimilares

Usuário-alvo

Sistema de Recomendação

Recomendação

Por que você gostou de:



Por que você não gostou de:

- **Como computar similaridade?**

Explicação por exemplos [Kim et al. '16]

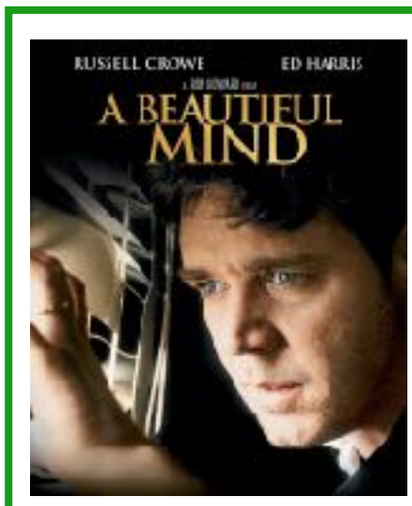
- **Seleção de exemplos**
  - similares
  - dissimilares

Usuário-alvo

Sistema de Recomendação

Recomendação

Por que você gostou de:



Por que você não gostou de:

- **Como computar similaridade?**

$$D\left(< \texttt{visual}_1 + \texttt{textual}_1 >, < \texttt{visual}_2 + \texttt{textual}_2 >\right)$$

Indução multimodal + *metric learning*

- **Limitação 1: encontrar vizinhança relevante**

  Métricas de distância dependente da modalidade

- **Aprendizado da vizinhança**
  - *Multimodal metric learning*

- **Limitação 2: explicar a previsão**
  - Explicação de **imagens**: **superpixels**

- **Explicação por exemplos**
  - *Similares e não-similares*

Indução multimodal + *metric learning*

- **Limitação 1: encontrar vizinhança relevante**

  Métricas de distância dependente da modalidade

Doutorado

- **Aprendizado da vizinhança**
  - *Multimodal metric learning*

- **Limitação 2: explicar a previsão**
  - Explicação de **imagens**: **superpixels**

Doutorado

- **Explicação por exemplos**
  - *Similares e não-similares*

Objetivo

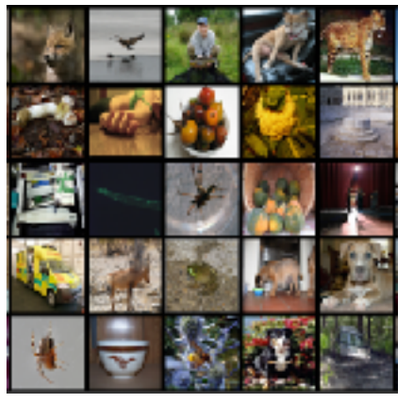
- Utilizar informação multimodal para:
  1. Melhorar a interpretabilidade das previsões
  2. *Few-shot learning*

Few-shot learning

Few-shot learning *[Sachin and Larochelle '16]*

- Imagens

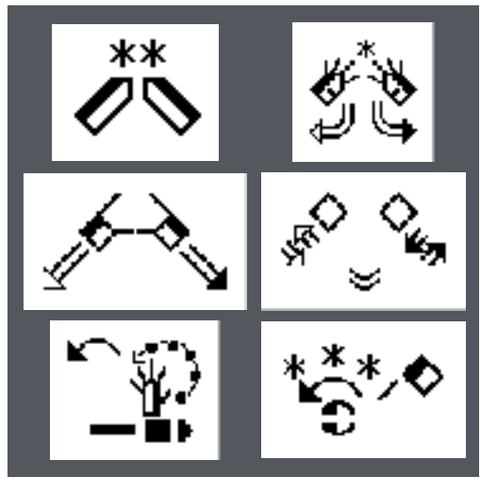
Imagenet



- *14M imagens*
- *27 classes*
- *+21k sub-classes*

VS

SignBank



- *~ 7K sinais*
- *~ 7K termos*

Few-shot learning *[Sachin and Larochelle '16]*

- Enriquecimento da base de dados
- Novas representações dos dados

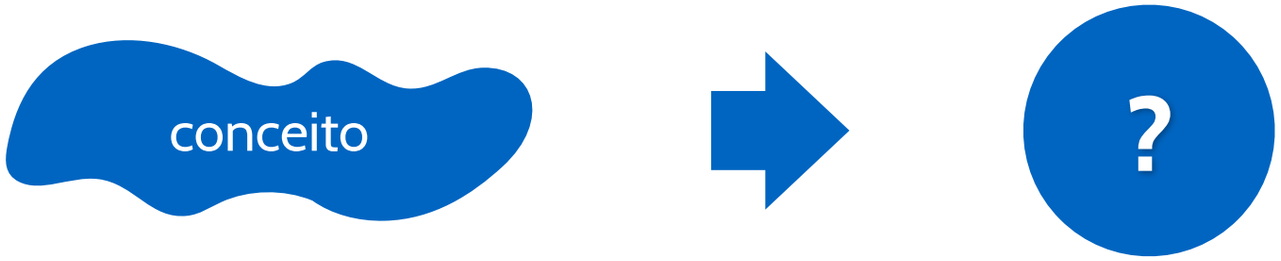
Domínio específico

Few-shot learning *[Sachin and Larochelle '16]*

- Enriquecimento da base de dados
- Novas representações dos dados

Domínio específico

- Criação automática de sinais para surdos



Criação automática de sinais para surdos

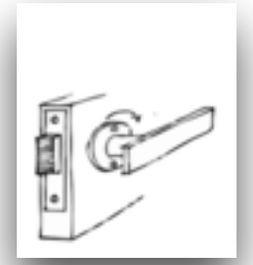
Criação de sinais [Souza et al. '18]

Criação de sinais [Souza et al. '18]

- Foco em educação

Sala 1

fechadura

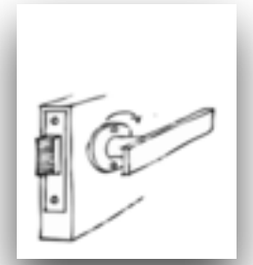


Criação de sinais [Souza et al. '18]

- Foco em educação

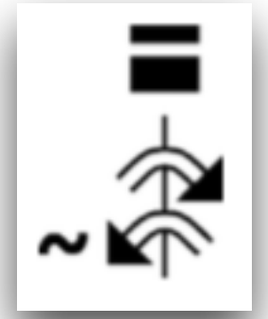
Sala 1

fechadura



Acordo:
aluno$_1$ / intérprete$_1$

sinal$_1$

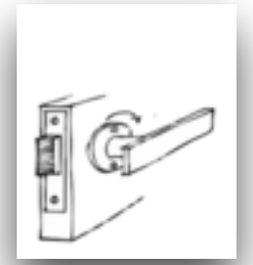


Criação de sinais [Souza et al. '18]

- Foco em educação

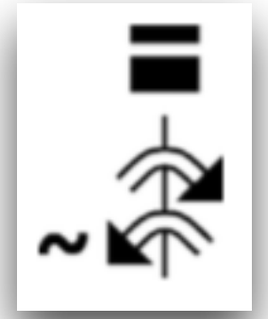
Sala 1

fechadura



Acordo:
aluno$_1$ / intérprete$_1$

sinal$_1$



Sala 2

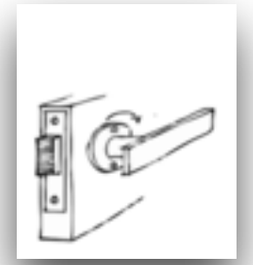
fechadura

Criação de sinais [Souza et al. '18]

- Foco em educação

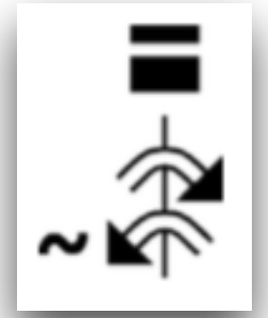
Sala 1

fechadura



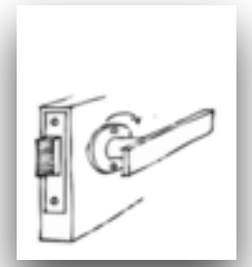
Acordo:
aluno$_1$ / intérprete$_1$

sinal$_1$



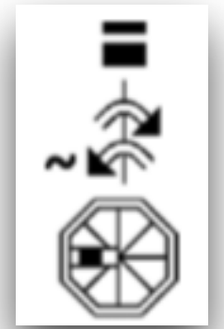
Sala 2

fechadura



Acordo:
aluno$_2$ / intérprete$_2$

sinal$_2$



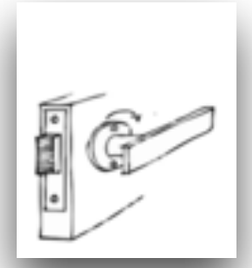
Revisão da literatura

Criação de sinais [Souza et al. '18]

- **Sinais técnicos**
  - Desenho arquitetônico

compasso

fechadura

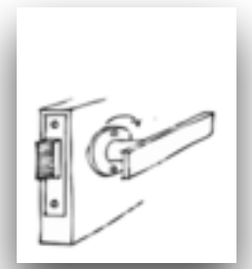


Criação de sinais [Souza et al. '18]

- **Sinais técnicos**
  - Desenho arquitetônico

compasso

fechadura



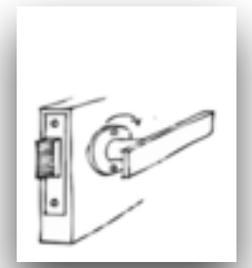
- **Libras**
  - 2a língua oficial do Brasil

Criação de sinais [Souza et al. '18]

- **Sinais técnicos**
  - Desenho arquitetônico

compasso

fechadura



- **Libras**
  - 2a língua oficial do Brasil

- ***SignWriting***
  - Valerie Sutton (1974)
  - Mãos, movimentos, expressões faciais e pontos de contato

casa

Criação de sinais [Souza et al. '18]

1 Descrição → 2 Ordenação de candidatos → 3 Junção de ordenações → 4 Símbolo candidato → 5 Composição

Conceito: **casa**

1

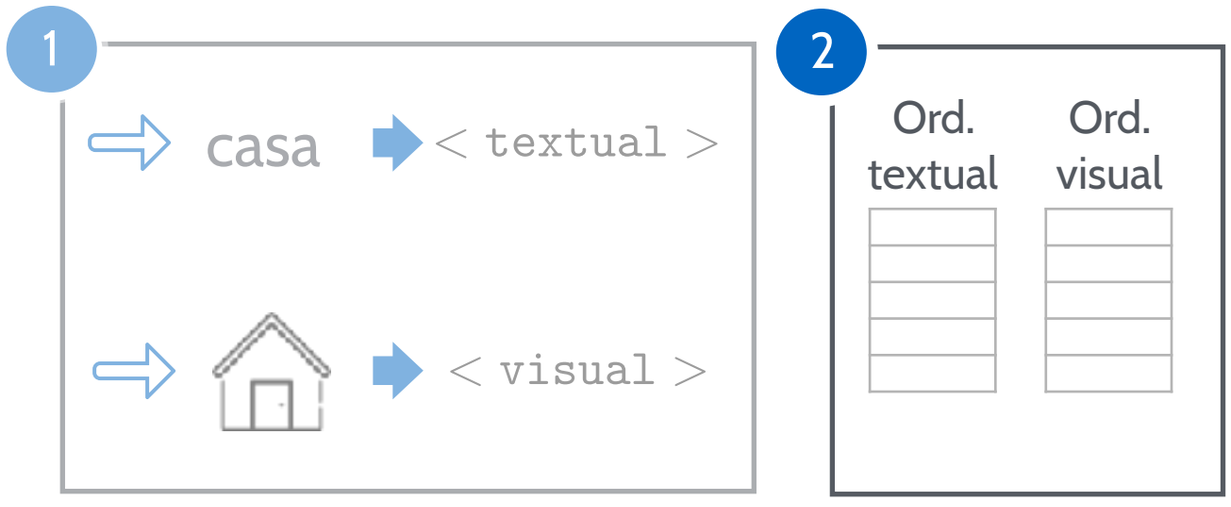
⇨ casa ➡ < textual >

⇨  ➡ < visual >

Criação de sinais [Souza et al. '18]

1. Descrição
2. **Ordenação de candidatos**
3. Junção de ordenações
4. Símbolo candidato
5. Composição

Conceito: **casa**

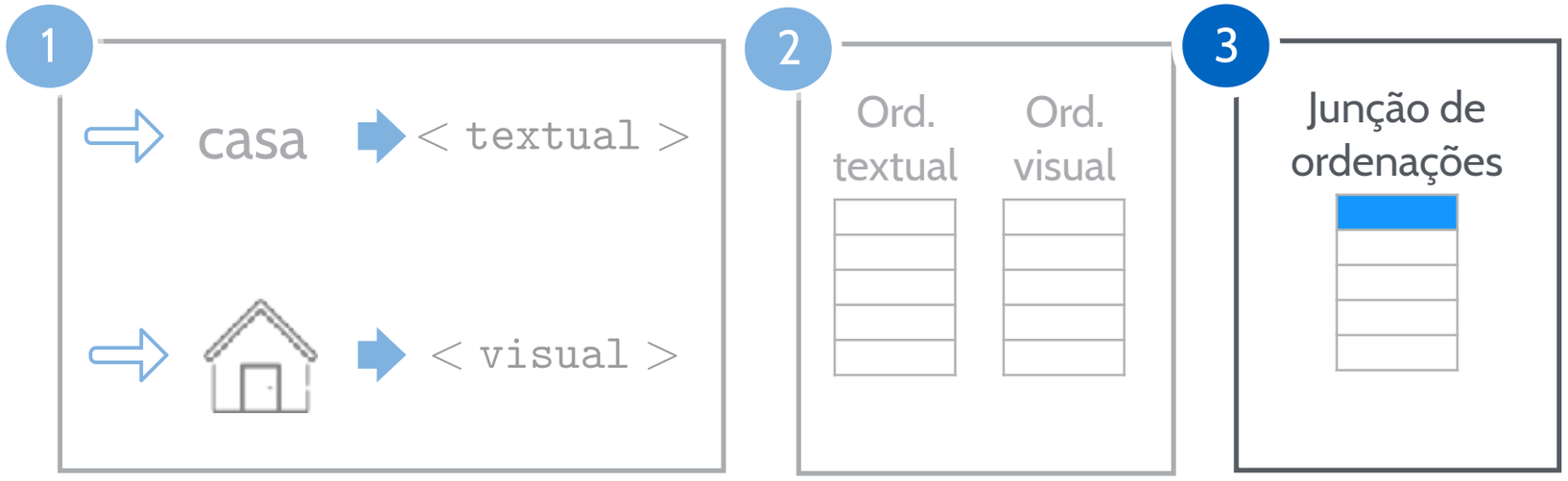


- **Levenshtein**

Criação de sinais [Souza et al. '18]

1 Descrição → 2 Ordenação de candidatos → 3 **Junção de ordenações** → 4 Símbolo candidato → 5 Composição

Conceito: **casa**

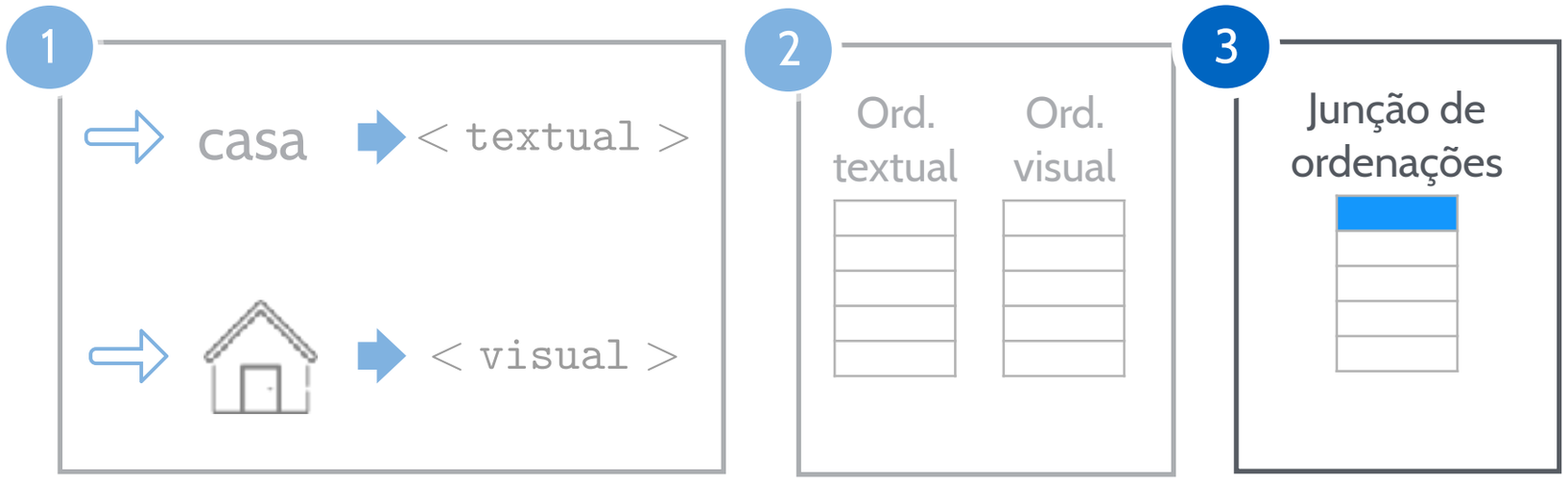


$$S_j^m = \left(1\left[S_j^t > \varepsilon_t\right] \times 1\left[S_j^i > \varepsilon_i\right]\right) \times \left(S_j^t + S_j^i\right)$$

Criação de sinais [Souza et al. '18]

1 Descrição → 2 Ordenação de candidatos → 3 **Junção de ordenações** → 4 Símbolo candidato → 5 Composição

Conceito: **casa**

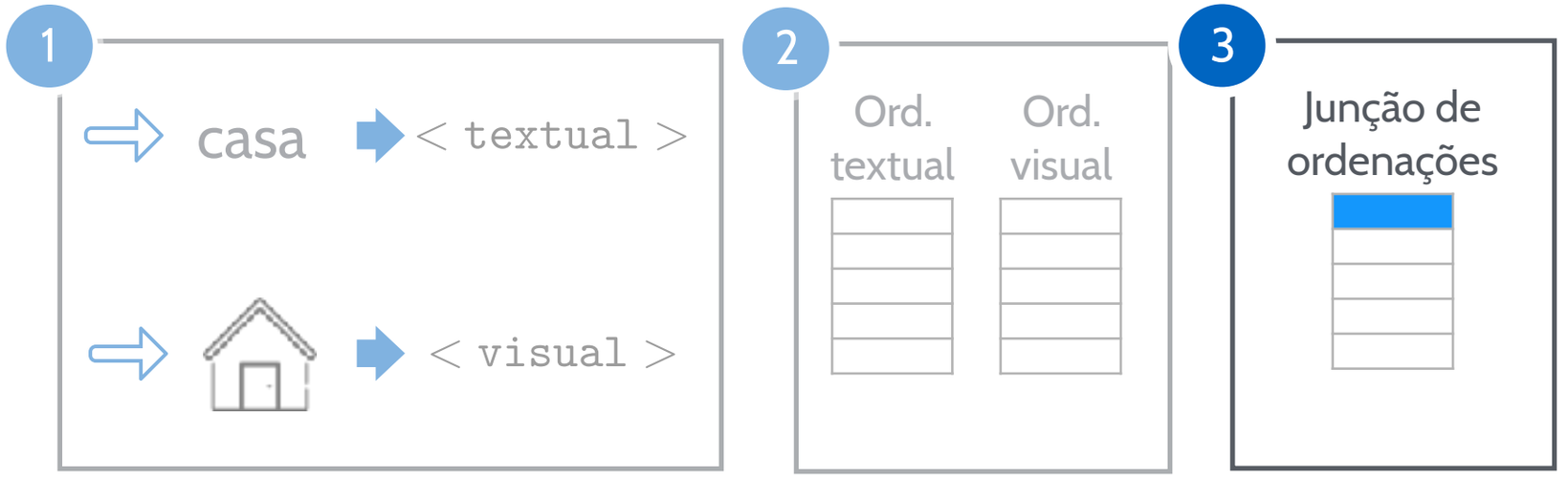


$$S_j^m = \left(1\left[S_j^t > \varepsilon_t\right] \times 1\left[S_j^i > \varepsilon_i\right]\right) \times \left(S_j^t + S_j^i\right)$$

Criação de sinais [Souza et al. '18]

1 Descrição → 2 Ordenação de candidatos → 3 **Junção de ordenações** → 4 Símbolo candidato → 5 Composição

Conceito: **casa**

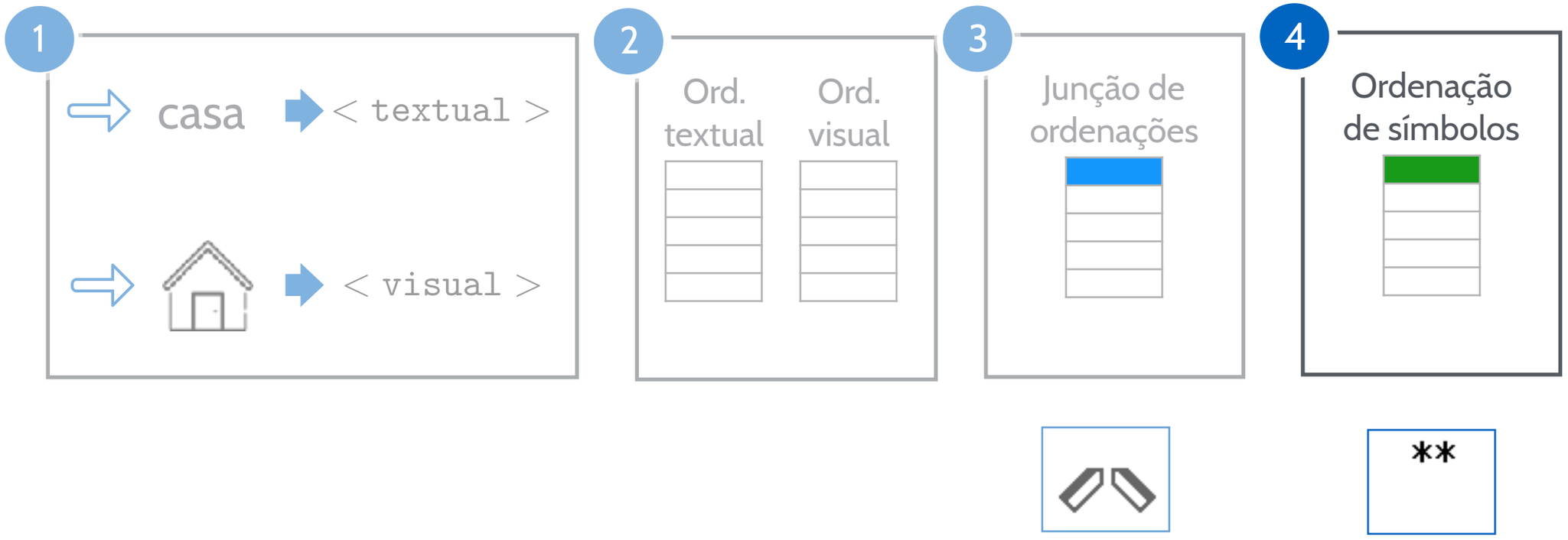


$$S_j^m = \left(1\left[S_j^t > \varepsilon_t\right] \times 1\left[S_j^i > \varepsilon_i\right]\right) \times \left(S_j^t + S_j^i\right)$$

Criação de sinais [Souza et al. '18]

1 Descrição → 2 Ordenação de candidatos → 3 Junção de ordenações → 4 **Símbolo candidato** → 5 Composição

Conceito: **casa**



Criação de sinais [Souza et al. '18]

1. Descrição
2. Ordenação de candidatos
3. Junção de ordenações
4. Símbolo candidato
5. **Composição**

Conceito: **casa**

Criação de sinais [Souza et al. '18]

- **Sinais técnicos**
  - Desenho arquitetônico

- **Abordagem multimodal**
  - Texto (termos) + imagens (ilustrações)

- **Validação com a comunidade surda**
  - *SignWriting* + desenho arquitetônico

Publicação	Autores	Trabalho	Evento
Periódico	Souza et. al.	A Computational Approach to Support the Creation of Terminological Neologisms in Sign Language	CAEE'18
Patente	Pádua et. al.	Sistema e Métodos para Geração, Preservação e Sinalização de Neologismos Terminológicos em Línguas de Sinais	-

Limitações

Limitações

- **Limitação 1:**

  Não existem ilustrações dos conceitos

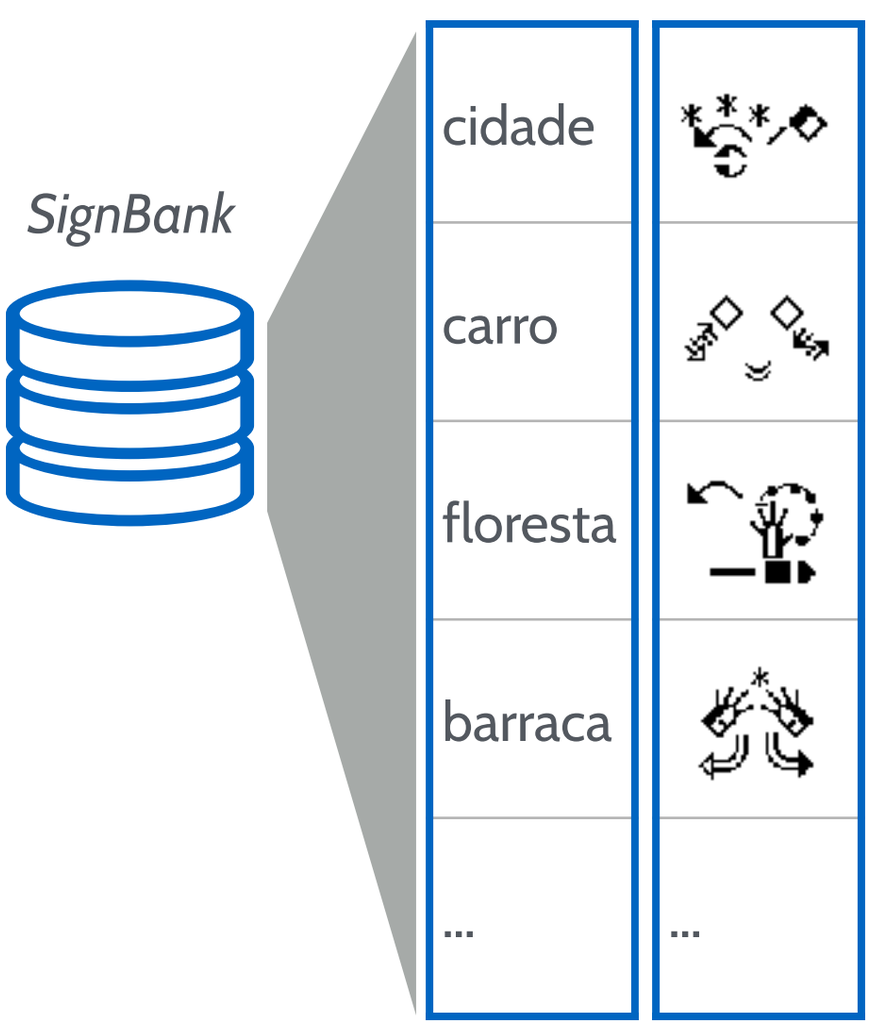
- **Limitação 2:**

  Seleção de sinais candidatos

Abordagem proposta

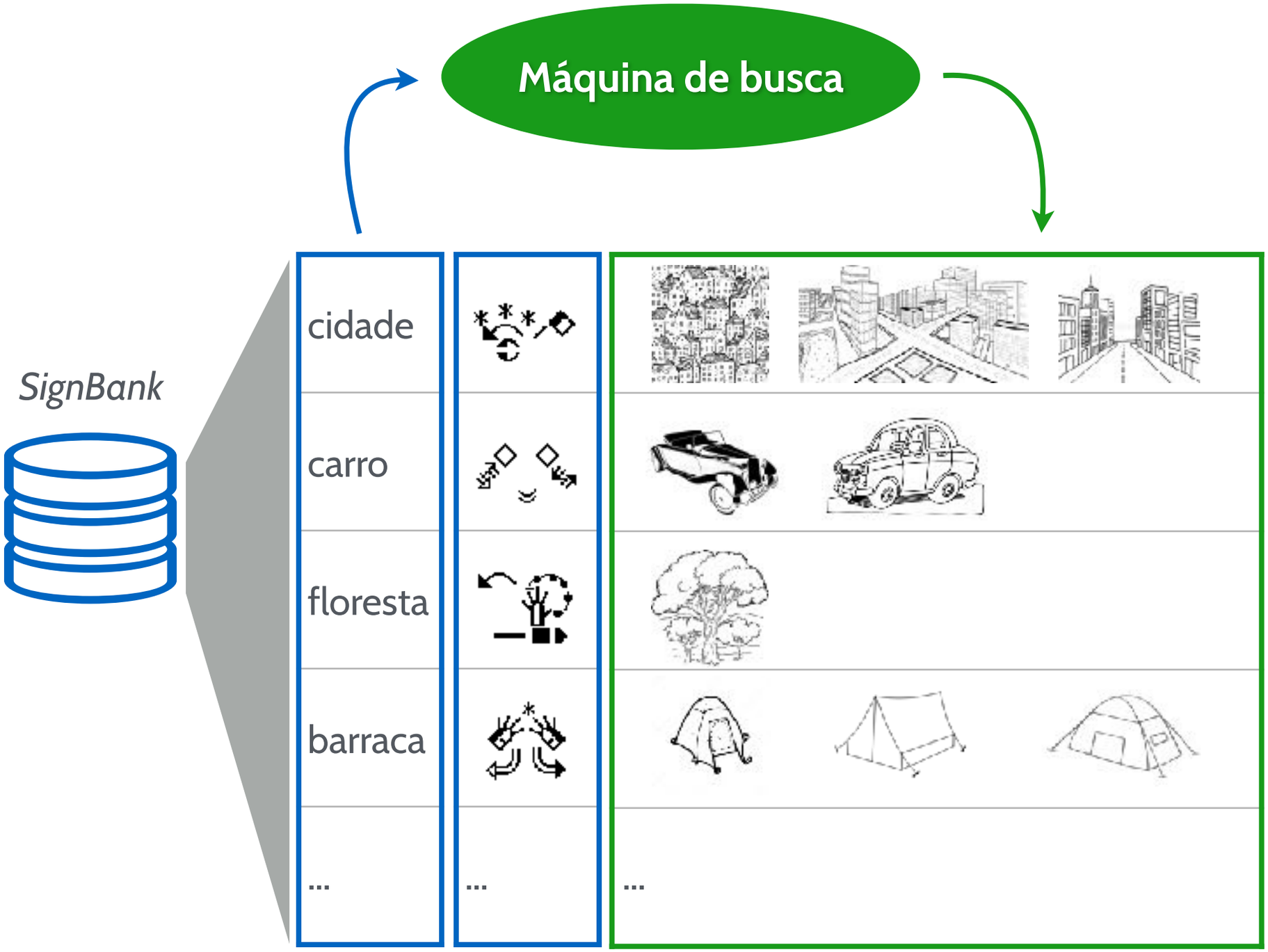
Como enriquecer base de imagens?

1. Novas imagens



Como enriquecer base de imagens?

1 Novas imagens



Como enriquecer base de imagens?

1. **Novas imagens**

- **Desafio 1:**
  Poucas palavras para consulta

- **Desafio 2:**
  Filtro dos resultados

Como selecionar sinais candidatos?

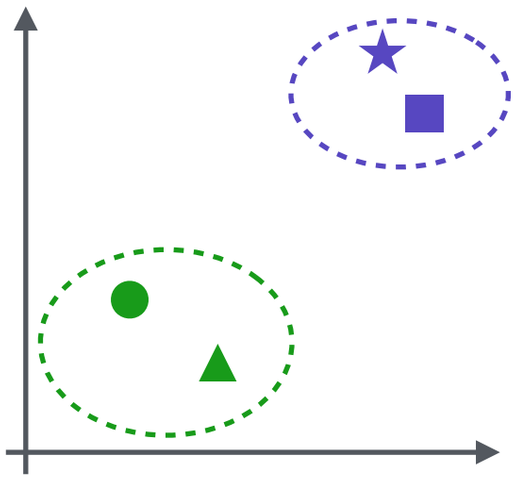
2 Agrupamento multimodal de sinais

Ord. textual

●	mala
▲	bolsa
★	viagem
■	turismo

Como selecionar sinais candidatos?

2 Agrupamento multimodal de sinais

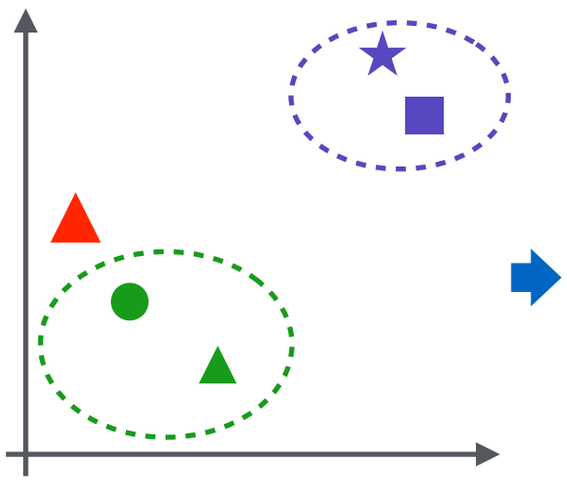


Ord. textual

mala	
bolsa	
viagem	
turismo	

Como selecionar sinais candidatos?

2 Agrupamento multimodal de sinais



Ord. textual

mala	0.7
bolsa	0.6
viagem	0.3
turismo	0.2

Como selecionar sinais candidatos?

2 Agrupamento multimodal de sinais

- **Fusão tardia**
  - Agrupamentos distintos para cada modalidade

- **Fusão antecipada**
  - *Multimodal metric learning*

Abordagem proposta

- **Limitação 1:**

  Não existem ilustrações dos conceitos

- **Recuperação de ilustrações descritivas em máquinas de busca**

- **Limitação 2:**

  Seleção de sinais candidatos

- **Agrupamento de sinais candidatos**
  - Fusão tardia
  - Fusão antecipada (*multimodal metric learning*)

Abordagem proposta

- **Limitação 1:**

  Não existem ilustrações dos conceitos

Mestrado

- **Recuperação de ilustrações descritivas em máquinas de busca**

- **Limitação 2:**

  Seleção de sinais candidatos

Mestrado

- **Agrupamento de sinais candidatos**
  - Fusão tardia
  - Fusão antecipada (*multimodal metric learning*)

Sumário

Sumário das contribuições

- **Interpretabilidade**
  - Algoritmos para representação de dados multimodais
  - Nova técnica de explicação de previsões
- ***Few-shot learning***
  - Nova base de dados de sinais (sinais + ilustrações)
  - Algoritmos de agrupamento para seleção de sinais candidatos

Referências

ML **[Xie et al. ’18]** Xie, Pengtao, Wei Wu, Yichen Zhu, and Eric P. Xing. "Orthogonality-Promoting Distance Metric Learning: Convex Relaxation and Theoretical Analysis–Supplements.” In *Proceedings of the International Machine Learning Conference*, ICML, 2018.

FS **[Wang et al. '18]** Wang, Yu-Xiong, Deva Ramanan, and Martial Hebert. "Learning to model the tail." In *Advances in Neural Information Processing Systems*, pp. 7029-7039. NIPS. 2017.

CS **[Souza et al. '18]** Souza, Celso L., Flávio LC Pádua, Vera LS Lima, Anisio Lacerda, and Carlos AG Carneiro. "A computational approach to support the creation of terminological neologisms in sign languages." *Computer Applications in Engineering Education*26, no. 3: 517-530. 2018.

IT **[Miller '17]** Miller, Tim. "Explanation in artificial intelligence: insights from the social sciences." *arXiv preprint arXiv:1706.07269* (2017).

IT **[Oliveira et al. ’16]** Oliveira, Samuel, Victor Diniz, Anisio Lacerda, and Gisele L. Pappa. "Evolutionary rank aggregation for recommender systems." In *IEEE Congress on Evolutionary Computation (CEC)*, pp. 255-262. IEEE, 2016.

IT **[Hendricks et al. '16]** Hendricks, Lisa Anne, Zeynep Akata, Marcus Rohrbach, Jeff Donahue, Bernt Schiele, and Trevor Darrell. "Generating visual explanations." In *European Conference on Computer Vision*, pp. 3-19. Springer, Cham, 2016.

Referências

FS **[Triantafillou et al. '17]** Triantafillou, Eleni, Richard Zemel, and Raquel Urtasun. "Few-shot learning through an information retrieval lens." In *Advances in Neural Information Processing Systems*, pp. 2255-2265. NIPS. 2017.

FS **[Sachin and Larochelle '16]** Ravi, Sachin, and Hugo Larochelle. "Optimization as a model for few-shot learning." (2016). In *International Conference on Learning Representations.* ICLR. 2016.

FS **[Vinyals et al. '16]** Vinyals, Oriol, Charles Blundell, Tim Lillicrap, and Daan Wierstra. "Matching networks for one shot learning." In *Advances in Neural Information Processing Systems*, pp. 3630-3638. NIPS. 2016.

IT **[Ribeiro et. al. '16]** Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should i trust you?: Explaining the predictions of any classifier." In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144. ACM, 2016.

IT **[Kim et. al. '16]** Kim, Been, Rajiv Khanna, and Oluwasanmi O. Koyejo. "Examples are not enough, learn to criticize! criticism for interpretability." In *Advances in Neural Information Processing Systems*, pp. 2280-2288. NIPS. 2016.

IT **[Lake et al. '15]** Lake, Brenden M., Ruslan Salakhutdinov, and Joshua B. Tenenbaum. "Human-level concept learning through probabilistic program induction." *Science* 350, no. 6266 (2015): 1332-1338.

Referências

ML

[Xie '15] Xie, Pengtao. "Learning compact and effective distance metrics with diversity regularization." In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 610-624. Springer, Cham, 2015.

IT

[Cheng et al. '14] Cheng, Hui, Jingen Liu, Saad Ali, Omar Javed, Qian Yu, Amir Tamrakar, Ajay Divakaran et al. "Multimedia event detection and recounting." *SRI-Sarnoff AURORA at TRECVID* (2014).

IT

[Letham et al. '15] Letham, Benjamin, Cynthia Rudin, Tyler H. McCormick, and David Madigan. "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model." *The Annals of Applied Statistics* 9, no. 3 (2015): 1350-1371.

ML

[Xie and Xing '13] Xie, Pengtao, and Eric P. Xing. "Multi-modal distance metric learning." In Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, pp. 1806-1812. AAAI Press, 2013.

IT

[Lacerda et al. '06] Lacerda, Anisio, Marco Cristo, Marcos André Gonçalves, Weiguo Fan, Nivio Ziviani, and Berthier Ribeiro-Neto. "Learning to advertise." In *Proceedings of the 29th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 549-556. ACM, 2006.

ML

[Xing et. al '03] Xing, Eric P., Michael I. Jordan, Stuart J. Russell, and Andrew Y. Ng. "Distance metric learning with application to clustering with side-information." In *Advances in neural information processing systems*, pp. 521-528. NIPS. 2003.

Árvore de decisão